The development of a British assessment of inferential and idiom comprehension for 5:00 to 9:11 year-old typically-developing children

Anne Hewitt

A thesis submitted in partial fulfilment of the requirements of Manchester Metropolitan University for the degree of Doctor of Philosophy

Department of Health Professions

2017

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following people who have helped me complete this PhD thesis.

My supervisors Jois Stansfield, Juliet Goldbart and Kevin Rowley.

The MMU students who helped me with some of the data collection.

Julie Wright for agreeing to fund this PhD from her departmental budget.

The schools and pupils who allowed me to work with them and who gave many amusing responses to my test questions.

My family and work colleagues for their support during my studies.

TABLE OF CONTENTS

CHAPTER TWO:

LIST OF TABLES

PAGE

LIST OF FIGURES
PAGE

Abstract

Background

Understanding of inferences and non-literal language, such as idioms, is critical for successful communication and academic learning. Assessment of inferential and idiom comprehension is essential for children who are showing difficulties in these areas so that appropriate intervention and support can be provided. There is very little information on the typical development of these specific areas of verbal comprehension in the literature and there are very few current assessments of inferential and idiom comprehension for British school-aged children. While many assessments that do exist have face validity, very few are standardised. Some children with comprehension difficulties do well on existing picture-based assessments of verbal comprehension but they demonstrate significant difficulties with more abstract language comprehension. There is a gap in the current battery of assessments available to paediatric speech and language therapists for assessing inferential and idiom comprehension in detail.

Aims

The primary aims of this thesis were to develop a robust standardised British assessment of inferential and idiom comprehension for 5:00 to 9:11 year-old children, to provide supporting validity and reliability data for the newly devised assessment, to provide normative and statistically significant data for inferential and idiom comprehension in typically developing children aged 5:00 to 9:11, to provide qualitative information on the typical development of these areas of verbal comprehension and to carry out exploratory studies using the new assessment with children with communication impairments. Secondary aims of the study were to examine if there was any relationship between gender and test performance and between socio-economic status and test performance.

Methods & Procedures

A new assessment, the Hewitt Inferential Comprehension and Idioms Test (HICIT) was created following a review of the literature and of the existing assessments in these areas of verbal comprehension. A pilot study was carried out with sixty-two 4:06 to 10:06 year-old children in two primary schools in the North West of England. The fourteen section, 210 item pilot test was reduced to the twelve section, 140 item final version of the HICIT. This was then trialled with a further 200

children, aged 5:00 to 9:11, making an overall standardisation sample of 250 children.

Normative data were provided from the application of descriptive and inferential statistics to the results. A two-way ANOVA examined the effects of age group and gender on test scores and a one-way ANOVA and post-hoc independent T tests looked at the relationship between socio-economic status and test performance.

Outcomes & Results

The construct validity of the test was examined with exploratory and confirmatory factor analyses. These demonstrated a single factor loading and good model fit measurements. The reliability of the test, as assessed by Cronbach's Alpha was moderate, similar to an existing published British standardised verbal language assessment. Rasch analysis indicated that the internal consistency of the test was good. The inter-rater reliability of 98.6% was excellent.

The descriptive and inferential statistics demonstrated that there was a developmental progression between the age groups for inferential and idiom comprehension but that there was no effect for gender. The idioms sub-section was the only section not to reach ceiling scores by 9:11.

The results for the relationship between socio-economic status and test scores were inconclusive.

Conclusions & Implications

Quantitative analysis of the HICIT data demonstrated that the test is a robust assessment of inferential and idiom comprehension. Some sub-tests of the test are more robust with different age groups so different versions of the test could be used with different age groups.

Qualitative analysis of the test responses, exploratory case studies using the HICIT to assess four children with communication impairments and feedback from practising speech and language therapists produced useful information on the possible applications of the test and suggested that it could be very useful to assist in differential diagnosis of different types of communication impairments.

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

This chapter begins with a brief justification for the development of a test of inferential and idiom comprehension and is then is divided into the following eight sections:

1. Definitions of the terminology used in this study.
2. The clinical rationale for developing an assessment of inferential and idiom comprehension.
3. Inferential comprehension in typically developing children and children with communication impairments.
4. Idiom comprehension in typically developing children and children with communication impairments.
5. Social inferencing in typically developing children and children with communication impairments.
6. Creating a language test.
7. Current formal and informal assessments of inferential and idiom comprehension.
8. The chapter summary containing the study aims and hypotheses.

Justification for the development of a test of inferential and idiom comprehension

The current researcher is one of those clinicians "for whom research functions not as a distraction from practice but as a development of it." (Schon, 1983, pii). Inferential and idiom comprehension are areas of language that come under the umbrella of verbal comprehension and pragmatics (the understanding and use of language in social contexts). The need to develop a British assessment of inferential and idiom comprehension evolved during the current researcher's twenty five years' experience as a paediatric Speech and Language Therapist (SLT) working with children with a range of communication impairments. Her first NHS SLT job was working in a language unit attached to a mainstream school. This included children who were described as having a 'semantic pragmatic language disorder' (SPD). She later developed a specialism in children with Autism Spectrum Disorders (ASD) and noticed that there was a large overlap of features between the children with SPD and those with high functioning ASD. These children were able to perform within normal limits or were mildly delayed on tests of comprehension of grammatical structures and receptive vocabulary, but they demonstrated severe difficulties in other areas of verbal comprehension, particularly inferential and idiom comprehension. Her SLT colleagues have also frequently encountered this pattern of difficulties and have expressed their frustration that there is currently no comprehensive standardised assessment of inferential and idiom comprehension on the market. The development of such a tool is therefore crucial for accurate assessment and intervention planning for children with comprehension impairments.

Before moving on to the rationale from the literature for developing this assessment, definitions of comprehension and communication impairments will be given.

CHAPTER ONE SECTION ONE: 1.1 Definitions of the terminology used in this study

1.1.1 Definitions of verbal comprehension and its sub-components

A variety of terms are used in the literature to describe verbal comprehension skills, some of which are interchangeable. Speech and Language Therapists use the terms ‘verbal reasoning’ and ‘high-level comprehension’ to refer to abstract language comprehension. ‘Inferential comprehension’ and ‘inductive reasoning’ are also interchangeable terms. The existence of multiple definitions in this area demonstrates the complexity and interconnectivity of the sub-components of verbal communication. Verbal reasoning, the understanding of figurative expressions, the use of language in social settings, lexical skills (word knowledge, word-finding and word definition), the production of complex syntax, and discourse proficiency are all intricately related aspects of later language development (Nippold, 1998). The definitions given below will be the ones adopted in this thesis.

1.1.2 Definition of comprehension

Comprehension refers to understanding. Verbal comprehension is the understanding of spoken language. It is also called receptive language.

Leinonen and Letts (1997) distinguish between linguistic comprehension (understanding of syntax and semantics) and pragmatic comprehension (where meaning has to be worked out from linguistic and contextual cues).

1.1.3 Definition of pragmatics

Dennis et al (2001) define pragmatics as the domain of language concerned with how speakers use language to effect successful communication. They also emphasize that inferencing is a key component of pragmatics.
Letts and Leinonen (2001) describe pragmatics as the use and understanding of language in context and they state that it is the interface between language and cognition. They suggest that cognitive rather than linguistic deficits could be the cause of pragmatic impairments. They highlight that most studies of pragmatic impairments mention verbal comprehension problems and state that pragmatic comprehension difficulties include problems with verbal reasoning, making inferences about actions, and an inflexible style of making inferences.

1.1.4 Definition of metapragmatics

This is an awareness in the speaker of the intention and effects of their pragmatic skills in communication.

1.1.5 Definition of verbal reasoning

Verbal reasoning is understanding and reasoning using spoken and written language. It involves listening, reading and the ability to think constructively. It includes the ability to interpret meaning from given and implied information and complex thinking tasks, such as analysis, synthesis and evaluation of text (Deary et al, 2007, in Spencer et al, 2016). Verbal reasoning is a component of cognitive ability, and it encompasses almost all learning in education. However, it cannot be defined by a single coherent cognitive theory as it involves many different thought processes. Reasoning is a form of thinking in which conclusions are drawn from a set of facts. Inductive reasoning is where general conclusions are drawn from particular facts or scenarios which might or might not be true. Nippold (1998) describes inductive problems as those where the given information supports the solution but is not sufficient evidence for it on its own. Inductive problems include: analogies (eg bear is to cub as cow is to ____); proportions (eg 1 package is to 2 cans as 3 packages is to_____ cans); series completions (eg June, August, October ______); classifications (eg odd one out from: aunt, cousin, sister, friend); understanding figurative language such as metaphors (‘She is a walking dictionary’) and proverbs (‘Two wrongs don’t make a right’).

1.1.6 Definition of high level language

Rhea and Norbury (2012) define high level language as language that goes beyond grammar and vocabulary which includes inferencing, reasoning and non-literal language comprehension. This is the definition of high level language that will be adopted in this thesis.

1.1.7 Definition of idioms

Crystal (2008) defines idiom as a term used in grammar and lexicology to refer to a sequence of words which is semantically and syntactically restricted so that it functions as a single unit. The meaning of the overall utterance cannot be deduced from the individual words (eg ‘it’s raining cats and dogs’).

Cain et al (2009) define idioms as a form of figurative language that usually have both a literal and a figurative meaning, depending on the context. For example, the sentence ‘someone spilled the beans’ can describe someone tipping out beans from a tin (literal) or someone giving away a secret (figurative).

Both the literality and the compositionality (the way in which the literal meanings of the word constituents contribute to the overall figurative meaning) of idioms contribute to how easy or hard they are to understand (Caillies and Le Sourn-Bissaoui, 2008).

Metalinguistic (in particular metasemantic) competence is required to work out the meaning of unfamiliar figurative expressions such as metaphors, proverbs and idioms (Nippold, 1998). Theory of mind skills are also required to make the literal-figurative link with idioms and other figurative language. Norbury (2004) points out that theory of mind ability enables one to understand speaker intention, thus helping to signal some non-literal meanings. However, she emphasizes that the role of theory of mind in idiom comprehension has not yet been fully investigated.

1.1.8 Sub-types of idioms

Cain et al (2009) and most other researchers sub-categorise idioms into transparent (or decomposable, or semantically analyzable) idioms and opaque (or non-decomposable or semantically nonanalyzable) idioms.

Transparent idioms have a clear overlap between the literal and figurative meanings of the phrase (eg ‘to speak your mind’).

Opaque idioms cannot be broken down word by word and their meanings cannot be derived successfully by semantic analysis (eg ‘to bite the dust’).

1.1.9 Definition of inferential comprehension

Inferential comprehension is the ability to understand language in context and to fill in information that is not explicitly stated. In other words, the ability to ‘read between the lines’. To be able to do this successfully requires drawing on world knowledge, previous contextual experience and social scripts. Leinonen and Letts (1997) include inferential understanding under the umbrella of pragmatic comprehension as it involves going beyond linguistic comprehension to the integration of internal world knowledge, previous experiences and ‘social scripts’ (what is said in certain situations) to be able to work out the possible intended meanings of the speaker.

Trabasso and Magliano (1996) state that the function of verbal inferences is to 'fill in' information that is not explicitly provided and to form connections amongst events in a text or conversation in order to enable comprehension of the overall meaning of the scenario. Adams et al (2001) describe inferential understanding as an ‘above-sentence level’ comprehension ability. This includes understanding of suprasegmental features of spoken language.

Ryder et al (2008) define inferencing as the integration of contextual information (via a reasoning process) and pragmatic demands in language comprehension to work out an intended meaning. They call this ability ‘implicature’.

Filiatrault-Veilleux et al (2015) state that inferential comprehension requires the ability to understand each word, morpheme and sentence; have a knowledge of the world; take into account others’ knowledge, motivations and intentions; understand expected social behaviour; and to ‘read between the lines.’

1.1.10 Sub-types of inferencing

Cain and Oakhill (1999) and Harley (1995, in Botting and Adams, 2005) define three major types of inference involved in verbal comprehension. The first is logical or text-connecting inferences, where the relationships between words or referents can be deduced or where the child needs to integrate information explicitly mentioned in the text to link ideas in two sentences. Cain and Oakhill's (1999) example of a text-connecting inference is: 'Michael got the drink out of his bag. The orange juice was very refreshing.' The inference is that Michael got the orange juice out of his bag. The second is elaborative inferences, where information from world knowledge helps in script building. The third is bridging or gap-filling inferences, where new information is related to old or the children need to integrate their own general knowledge with information in the text to fill in details that are not explicitly stated. Cain and Oakhill's (1999) example of a gap filling inference is: 'The girl put on her swimming costume, but the water was too cold to paddle in so she built sandcastles instead.' The inference is that the girl is at the seaside.

Saladana and Frith (2007) give the following example of a bridging inference. *John had left his umbrella at home…… He got very wet.* The bridging inference is that *it rained*. They further defined bridging inferences as 'off-line' if the inferences were made retrospectively (eg multiple choice questions after a text had been read) or 'on-line' if inferencing ability was assessed during reading (by measuring response times). Response times for primed inference questions are typically quicker than response times for non-primed questions.

Dennis et al (2001) describe four different kinds of pragmatic inferences: those connecting words to real world knowledge or to grammatical form; those that maintain text coherence, those that elaborate a text through the use of figurative language, and those that invoke inferences about the mental state of others.

Puche-Navarro and Millan (2007) define 'inductive inferences' as the establishment of a relation based on two elements that lead to a new understanding and 'relational inferences' as the establishment of two relationships in order to establish a third relationship, which leads to a new understanding.

1.1.11 Inferential comprehension in reading

Van Kleeck (2008) defines inferencing in reading as when a reader goes beyond information that is directly provided in a text to fill in information needed to understand the scenario or to elaborate on the information given. There is evidence that it is difficulty with inferencing that causes problems with reading comprehension and not vice versa (Cain and Oakhill, 1999).

Definitions of Different Communication Disorders and other developmental Disorders that include communication impairments

Over two decades ago Bishop and Adams (1992) emphasized that children with communication impairments were a heterogeneous population. There was no general agreement at that time as to how communication disorders should be sub-classified. Nowadays there is more consensus but there are still some disagreements about overlapping conditions.

The main internationally accepted diagnostic classification systems currently used to identify childhood developmental disorders are the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) (APA, 2013) and its European equivalent, the International Statistical Classification of Diseases and Related Health Problems (ICD-10, second edition) (WHO, 2004). The communication disorders identified come in section three of the DSM-5 (entitled neurodevelopmental disorders) and in section five of the ICD-10 (entitled mental and behavioural disorders).

The DSM-5 includes definitions of Autism Spectrum Disorder (ASD), social communication (pragmatic) disorder and language disorder.

The process of updating the DSM-4 (APA, 1994) to the DSM-5 (APA, 2013) involved a comprehensive review and discussion of the scientific literature by a working party; a review by the DSM-5 task force, scientific review and clinical and public health committees; and evaluation by the APA's board of trustees. The latest version of the ICD (ICD-11) was due to be available in 2015 but has been delayed until 2017.

1.1.12 Definition of Autism Spectrum Disorder (ASD)

The most recent label for this condition is Autism Spectrum Condition (ASC) which is interchangeable with the term ASD. The current researcher uses the term ASD as this is the term that is used in the DSM-5 and in many research studies.

DSM-5 (APA, 2013) defines ASD as persistent social interaction and communication impairments in multiple situations along with restricted, repetitive behaviour and interests. The difficulties have to manifest in the early developmental period and have to cause significant impairment. The very detailed DSM-5 and ICD-10 criteria for diagnosing ASD are given in Appendix i.

1.1.13 Definitions of social (pragmatic) communication disorder

The DSM-5 (APA, 2013) defines social (pragmatic) communication disorder (315.39) as persistent difficulties in the social use of verbal and nonverbal communication as manifested by the following:

1. Deficits in using communication for social purposes.
2. Impairment of the ability to change communication to match context or the needs of the listener.
3. Difficulties following the rules for conversation and storytelling.
4. Difficulties understanding what is not explicitly stated (inferences, idioms, metaphor, humour, etc).

These children should have no restricted interests and repetitive behaviours; the difficulties should not be attributable to hearing or other sensory impairment, motor dysfunction, any other neurological condition or developmental delay; and the symptoms should not be better explained by ASD.

The DSM-5 team report that previous editions of the DSM did not provide an appropriate diagnosis for people with the above symptoms. They were labelled as having 'Pervasive Developmental Disorder Not Otherwise Specified (PDDNOS) which led to inconsistent treatment and services. They point out that research demonstrates that communication disorders respond to treatment, so identifying distinct communication problems are an important first step in getting people appropriate intervention.

However, this statement appears to suggest that the DSM 5 considers social (pragmatic) communication disorder as qualitatively different from ASD and that it responds more effectively to treatment than ASD. The clinical experience of the current researcher leads her to disagree with this assumption. As can be seen from the above criteria for ASD and social communication disorder (SCD) there is much more overlap in symptoms than there are differences between the two sub-groups. The main difference is that the children with SCD should not show any restricted interests or repetitive behaviours. However, these are behaviours that can emerge and disappear throughout a child's development and this makes any clear cut long-term diagnosis of SCD problematic.

In fact this sub-group of communication disorder has been a controversial one since communication disorders were first labelled and it continues to be so. Rapin (1987, in Bishop and Adams, 1992) identified a sub-group of children with specific language impairment (SLI) as having primary problems with language content and use (semantic and pragmatic difficulties), in contrast to the main difficulty of the children with SLI being with language form. Bishop and Rosenbloom (1987, in Bishop and Adams, 1992) labelled the impairment Semantic Pragmatic Disorder (SPD). Children with SPD had good literal understanding of language but very impaired non-literal comprehension. Bishop and Adams (1992) did acknowledge that most of the literature on SPD at that time was based on clinical impression, and the validity of a distinctive subgroup within the SLI population was open to question. Leinonen and Letts (1997) used the term Pragmatic Impairment to describe the same group of children and they proposed a further sub-group of children within this who had a primary 'pragmatic-inferential' impairment. Bishop (2000, in Adams et al, 2009) re-labelled the sub-group as children with pragmatic language impairment (CwPLI). She reported that difficulty with making verbal inferences in the presence of relatively strong expressive language was a key defining clinical feature of these children. They either failed to make inferences at all or made bizarre inferences. They also interpreted non-literal language literally, were not able to take account of the listener's perspective and they produced socially inappropriate or stereotyped responses in conversations. She acknowledged that many of these pragmatic language impairments were also observed in children with high functioning autism (HFA). When Norbury carried out her idiom comprehension study in 2004 she stated that PLI was not part of the conventional diagnostic framework at that time and she reported that it was still not clear if these children were a distinct group or not. Many researchers still consider PLI to be a separate category from HFA but many view them as one and the same (Shields et al, 1996, Boucher, 1998). Norbury (2004) described the difficulties of maintaining the diagnostic boundaries between PLI, High Functioning Autism and Asperger's syndrome. Instead she based her classification on levels of language and used the sub-categories: language impairment (LI), pragmatic impairment (PI), ASD plus LI (ASL), and ASD only (ASO).

Norbury and Bishop (2002) conclude that the differential diagnosis of PLI and HFA is problematic and remains controversial. They point out that Rapin's (1987) identification of semantic and pragmatic difficulties was intended as a descriptive rather than a categorical label, that could apply to children with PLI and children with HFA and those who fall somewhere between the two

<u>1.1.14 Definitions of language disorders</u>

The DSM-5 (APA, 2013) defines language disorder (315.39) as a persistent difficulty in the acquisition and use of language across modalities (spoken, written, sign) due to deficits in comprehension or production. These include reduced vocabulary (word knowledge and use), limited sentence structure (grammar), and impairments in discourse (use of language to describe a topic or series of events or have a conversation).

The exclusion criteria are that the difficulties are not attributable to hearing or other sensory impairment, motor dysfunction, any other neurological condition or developmental delay.

The ICD-10 (WHO, 2004) includes 2 mutually exclusive language disorders:

F80.1: Expressive language disorder

F80.2: Receptive language disorder

The term specific language impairment (SLI) is often used for any of the above sub-categories of language impairment.

It can be seen from these definitions that there is also some overlap with the features of language impairment and the features of PLI and ASD, in particular the deficits in comprehension and discourse. Bishop (1987, in Bishop and Adams, 1992) points out that the term comprehension problems can cover impairments in many underlying processes ranging from auditory perception, attention, vocabulary, grammatical processing, memory, inferential skills, and understanding of non-literal language. More recent developments in the language disorders terminological debate are summarised by Bishop et al (2016) as part of the CATALISE consortium. They conducted an international Delphi poll to reach an agreed consensus on terms for language impairments from a wide range of professionals who specialised in this area. The conclusion was that, henceforth, the term 'language disorder' should be used for children with severe and persistent problems and the term 'developmental language disorder' (DLD) should be used for children formerly labelled as having specific language impairment (SLI). The Royal College of Speech and Language Therapists (RCSLT) will be endorsing these changes in terminology sometime in 2017. However, at the time of writing, the term SLI was still being used in the literature, so that is the term that has been adopted in this thesis.

1.1.15 Definitions of other communication impairments

The DSM-5 (APA, 2013) includes 'speech sound disorder' (315.39) and the ICD-10 (WHO, 2004) 'specific speech articulation disorder' (F80.0). These occur when speech difficulties are persistent and substantially below developmental expectations. They often co-occur with language impairment.

The DSM-5 (APA, 2013) also includes 'childhood-onset fluency disorder' (315.35) and the ICD-10 (WHO, 2004) 'stuttering/stammering)' (F98.5) which is another common childhood communication impairment. Stammering is not commonly associated with language impairment.

The ICD-10 (WHO, 2004) includes all communication disorders except stammering in the sub-group of mental and behavioural disorders entitled 'disorders of psychological development'. Stammering comes under the sub-group of 'behavioural and emotional disorders with onset usually occurring in childhood and adolescence'.

1.1.16 Definition of Attention Deficit Hyperactivity Disorder (ADHD)

The DSM-5 (APA, 2013) defines ADHD as a specific pattern of behaviour exhibited in multiple settings (eg, school and home), that causes difficulties in social, educational, or work settings. Children must show at least six symptoms from either (or both) the inattention group of criteria and the hyperactivity and impulsivity criteria, while older adolescents and adults (over age 17 years) must present with five. These features include the failure to pay close attention to details; difficulty organising tasks and activities; excessive talking, fidgeting, or an inability to remain seated in appropriate situations. The symptoms must not occur exclusively during the course of schizophrenia or another psychotic disorder and must not be better explained by another mental disorder, such as a depressive or bipolar disorder, anxiety disorder, dissociative disorder, personality disorder, or substance intoxication or withdrawal The authors acknowledge that ADHD often co-occurs with ASD.

The previous sections have outlined the key terms to be used throughout this thesis; defined the many sub-types of communication impairment that exist; and highlighted some differences and overlaps in some of the diagnostic categories. This chapter now proceeds to critique the literature on inferential and idiom comprehension acquisition in typically developing children and the difficulties with inferential and idiom comprehension in children with communication impairments.

CHAPTER ONE SECTION 2: 1.2 Clinical Rationale for developing an assessment of inferential and idiom comprehension – justification from the literature

Bishop and Adams (1992) describe a sub-group of children with semantic and pragmatic language difficulties who do much better on multiple-choice comprehension tests than on non-literal language comprehension assessments. They warn that the former assessments alone do not give a comprehensive picture of a child's functional verbal understanding. Dennis et al (2001) point out that many children with high functioning autism (HFA) master many language skills to an age-appropriate level but still exhibit inferential comprehension problems.

Adams et al (2001) also describe children who show good sentence comprehension but fail to draw inferences at story or text level, leading to an impaired understanding of cause and effect, sequences of events, and character motivation. These children are also reported to give tangential or bizarre answers to some questions. Botting and Adams (2005) highlight that, as children with language impairments develop, they can learn to perform reasonably well on standardised semantic and comprehension tests whilst still having significant inferential comprehension, conversational and social interactional difficulties, areas that are very difficult to assess objectively. This therefore weakens the face validity of some formal language assessments over time. The profile of relatively good comprehension of grammatical structures and weak inferential comprehension predominates in high functioning children with Autism Spectrum Disorders (HFA) but is also evident in children with specific language impairment (SLI) and children with pragmatic language impairments (PLI) (Bishop and Adams 1992, Conti-Ramsden and Gunn, 1986, Leinonen and Letts,1997, McTear 1985, all in Adams et al, 2001). Adams et al (2001) emphasize that the difficulties of children with language impairments in processing stories are not solely due to poor memory or difficulties understanding individual sentences, but to the inability to integrate information across the story to get the 'big picture'. This is largely an inferencing skill. Norbury and Bishop (2002) and Adams et al (2009) highlight that although disproportionate difficulty with verbal inferencing is often cited as a specific difficulty for children with PLI and HFA, there is still little empirical support for this view. Ryder et al (2008) state that the comprehension difficulties underlying children's expressive pragmatic difficulties have been under-researched. Inferential comprehension is crucial not only for successful social communication but also for educational achievement, as it is a skill that is essential for understanding the school curriculum. Bottings and Adams (2005) report that inferential comprehension skills are closely linked to educational attainment.

There are no existing assessments that assess these areas of language in detail and there are no British norms available for the development of these skills. Letts and Leinonen (2001) emphasize that inferential comprehension is linguistically and cognitively very demanding so is better developed in older children but point out that a clear developmental progression in the ability to manipulate inferential meaning has not yet been established. Adams et al (2009) state that research studies have shown clear developmental trends in inferencing ability from early childhood up to fifteen years of age in children with typical language development, but no clinical norms are available in this area. They also highlight that there are no existing methods for analysing the types of errors that children tend to make in inferential tasks.

Adams (2002) explains that it is very difficult to establish norms for pragmatic development in children due to the complex interaction of social, linguistic, cognitive and cultural influences. She also recommends that further research is required to establish clearer norms for the development of pragmatic ability, particularly for understanding of inference. Adams et al (2009) state that although clinicians have observed specific difficulty with verbal inference amongst children with language impairments, there is no empirical support for this. They emphasize that verbal inference is a very elusive behaviour to control in experiments and that many language impaired children may score at ceiling on simple inference tasks but present with substantial problems of inference comprehension in the classroom. They report that, so far, systematic studies of verbal inference in children with pragmatic language impairment have failed to confirm the clinical observation that they have specific difficulties in this area. This is not surprising, as the assessment of inference is very complex, particularly when it is assessed auditorily. Inferences can be assessed in different ways: online, offline and with or without nonverbal support. It is virtually impossible to control for all variables that could contribute towards inferential comprehension. For example: the child’s world knowledge, cultural experience and social-emotional knowledge (to be able to judge characters’ motives); the child’s understanding of grammatical structures; the abstractness of vocabulary used in the test; working memory demands; the sub-type of inference tested; and the distance of the implied inference from the initial context (Adams et al, 2009). Many research studies use sentence comprehension matched controls rather than age controls so language impaired children are being compared to children with a similar language level to get around the problem of inferential comprehension being assessed verbally.

Law (in Parsons and Branagan, 2005) states that there are many studies showing that SLT assessment and intervention are effective in developing the communication skills of pre-school children but more studies of the effectiveness of assessment and therapy for older children are urgently needed. Nippold (1998) highlights the fact that there is much more research into early language development than later language development. This is partly because most of the basic language development is complete by school age. However, Filiatrault-Veilleux et al (2015) state that more research is needed into inferential comprehension even in this age group as there is no detailed information on the way in which inferential comprehension develops in pre-school children (up to 6 years). More complex language continues to develop into adulthood but differs from early language development in terms of speed, salience, and substance.
Consequently, there are many more language assessments for younger than older children. In addition, many of the standardised tests that assess the language of school-age children and adolescents are based on clinical intuition and guess-work as there are no solid normative data for later language development (Nippold, 1998). The solution to this problem is further research into language development in the school-age and adolescent years and the design of good quality standardised and non-standardised measures that assess important aspects of later language development. There is a considerable body of evidence to show the link between the continued growth in language and the academic and social success of school-age children and adolescents. This means that reliable assessments of language skills in these age groups are essential in order to lead to appropriate intervention.

The majority of standardised assessments of the language of older children are American. The only published British assessments are the British Picture Vocabulary Scales (Dunn et al, 1998) which is a test of receptive vocabulary; the Test for Reception of Grammar (TROG-2) (Bishop, 2003); and the Assessment of Comprehension and Expression (ACE) 6-11 (Adams et al, 2001). Some American assessments such as the Clinical Evaluation of Language Fundamentals (CELF-4) (Semel et al, 2006) have had their vocabulary anglicised and been standardised on a British population. The TROG-2 is a test of understanding of grammatical structures only. The ACE includes a core test sub-section assessing inferential comprehension from a picture scenario and an extension sub-test testing non-literal language comprehension (idioms). It is therefore, only the ACE that includes normative data on the inferential and idiom comprehension of British children. Consequently, there is a need for a more detailed British assessment in these areas, which is the aim of this PhD thesis.

Adams et al (2009) conclude that we have not yet found a comprehensive clinical method of assessing verbal inference. They suggest that an assessment incorporating empathic and motivation-based inferences (including mental state verb type stimuli) could provide more useful results as these types of difficulties are well documented in children with ASD and PLI. They also recommend including an analysis of inferential comprehension error types for different sub-groups of children with communication impairments.
Inferential comprehension is being increasingly targeted in Speech and Language Therapy but, until we have more information on the typical development in this area and an effective assessment, we cannot properly target intervention or measure its efficacy (Bishop et al, 2009, Filiatrault-Veilleux, 2015). Ryder et al (2008) also state that the assessment of inferential comprehension is vital for providing appropriate educational support.

<u>CHAPTER ONE SECTION THREE: 1.3 Inferential comprehension in typically-developing children and children with communication impairments</u>

Each sub-section begins with a chart summarising the research studies in the particular area covered. Each study is then described in detail below its corresponding chart. The studies are critically evaluated together at the end of the whole section.

Table 1.1 Summary of the literature reviewed for inferential comprehension in typically-developing (TD) children

Researcher and date	**Type of inferential comprehension investigated and key findings**
Piaget (1965)	Analogical reasoning in the 4 stages of cognitive development.
Nippold (1998)	Analogical reasoning in TD children.
Leinonen and Letts (1997)	Inferential comprehension in TD children. 6 year-old children could answer inference questions.
Lynch and van den Broek (2007)	TD 6 year-olds could work out characters' goals in stories.
Bernicot et al (2007)	Implicature development in TD 6, 8, and 10 year-olds using a story-completion computer game.
Silva and Cain (2015)	Vocabulary knowledge was the sole predictor of literal and inferential comprehension in TD 4-6 year-olds.
Filiatrault-Veilleux et al (2015)	TD 3 year-olds understood internal responses; TD 4 year-olds understood characters' goals and problems; TD 5-6 year-olds understood consequence and solutions; TD 6 year-olds could predict what would come next.
Freed and Cain (2016)	TD 9-10 year-olds were better at answering global and local coherence inferences than TD 7-8 year-olds.

1.3.1 Developmental stages of analogical (or inductive) reasoning

Nippold (1998, p59) states that: "analogical reasoning occurs when an individual perceives similarities and differences between objects or events and uses that information to solve problems or to learn about the world." It is a basic human language and cognitive ability that begins in infancy, improves most in the school age and adolescent years, and continues to develop into old age. Verbal analogical reasoning is a mental construct where cognition and language converge. A child requires competence in both areas to be able to solve analogical reasoning problems. Both word knowledge and world knowledge are crucial for the development of verbal analogical reasoning.

Jean Piaget (1965), a French psychologist working at the Binet institute, was the first researcher to make a systematic study of children's cognitive development starting in the 1920s. From detailed observations of children's behaviour whilst learning and from the implementation of a series of ingenious cognitive tests he was able to produce a comprehensive theory of children's cognitive development. The three basic components to this theory are: schemas (which are the building blocks of knowledge), adaptation processes that enable the transition from one stage to another (these are equilibrium, assimilation and accommodation), and the identification of four specific stages of development (sensorimotor, preoperational, concrete operational and formal operational). In terms of analogical reasoning he demonstrated that 'pre-operational' children (aged 5-6) could sort pictures into pairs of things that go together. 'Concrete operational' children (aged 7-10) could sort two by two matrices where all four pictures go together, but sometimes needed to use trial and error strategies. 'Formal operational' children (aged 11-12) could do the two by two matrices without using trial and error strategies and they also gave superior explanations for their matching reasons than the younger children.

Nippold (1998) reports on a series of studies in the development of analogical reasoning. Overall findings are performance on analogical reasoning tests steadily increases with age; abstract analogies are more difficult to solve than concrete analogies; younger children are more easily deceived by plausible foils; and higher cognitive, general language and reading ability and a more reflective problem-solving style correlate to higher analogous reasoning ability.

A series of studies that test inferential and idiom comprehension in typically developing and communication impaired children will now be outlined. They will be critically evaluated at the end of each section, as there are recurring methodological issues in many of the studies.

1.3.2 Developmental stages of verbal inferential comprehension in typically-developing children

Botting and Adams (2005) define inferencing as the abstraction of information that is not explicitly presented. They propose that good inferencing skills require: good comprehension of presented material, the ability to meta-represent (grasp more than one concept at the same time) and the ability to take someone else's perspective (including understanding their moods and behaviours). They also state that inferential comprehension begins as young as 2 years of age and continues up until the age of 15.

Letts and Leinonen (2001) propose that the ability to answer linguistically and cognitively complex questions (such as inference) develops as children get older. However, they stress that a clear developmental progression in inferential comprehension has yet to be established and that comprehension of language in real communicative contexts is a complex process that continues to develop into adolescence. They highlight that different question types pose different demands on comprehension. For example, developmentally, 'What's that?' questions are easier to understand than 'Why?' and 'How?' questions as the latter require understanding of cause and effect (a cognitive skill). In addition, inferential comprehension questions vary in terms of how much non-linguistic context and general knowledge is needed to answer them correctly.

Milosky (1992, in Leinonen and Letts 1997) describe how children's comprehension develops by increasingly sophisticated use of context and acquired knowledge. Leinonen and Letts (1997) outline how children's ability to combine information to work out meaning (ie to make inferences) develops as they become more able to make connections on the basis of more subtle clues (both grammatical and pragmatic) and beyond the immediate context. They found that most 6 year-old children in their study were already able to answer many questions involving inference.

Lynch and van den Broek (2007) found that children as young as 6 were able to perform on-line 'think-aloud' inferencing tasks that involved working out the characters' goals in stories.

Bernicot et al (2007) assessed the developmental order of nonliteral comprehension (including implicatures) of 6, 8 and 10 year-old children in a story-completion task computer game. The results demonstrated that nonliteral language comprehension and metapragmatic knowledge were acquired in different orders. For comprehension, the developmental order of acquisition was: semantic-inference implicatures, indirect requests, idioms, sarcastic-inference implicatures. For metapragmatic knowledge, the order was: idioms, implicatures with a sarcastic or semantic inference, indirect requests. Researchers have established that difficulties making inferences in typically developing children are not solely due to limited verbal memory or world knowledge. Their inability to select the information relevant to making the inference was identified as the most likely source of difficulty (Oakhill and Cain, 2012).

Pike et al (2013) found that working memory was a significant factor in the development of inferential comprehension in school-aged children as it helped them to link relevant prior information with newly presented information.

Silva and Cain (2015) tested 82 typically developing 4-6 year-old children on literal and inferential questions linked to a picture book. The children's performance on both types of questions correlated with grammatical and receptive vocabulary knowledge. The inferential questions also correlated with verbal working memory. However, further statistical analysis demonstrated that, once age and cognitive level were taken into account, vocabulary knowledge was the sole predictor of both literal and inferential comprehension.

Filiatrault-Veilleux et al (2015) did a scoping review of the literature on the development of inferential comprehension, specifically causal inferences in story grammar in 3-6 year-old children.

They outline the structure of a story and corresponding inferential questions as:

1. A setting
2. A triggering event: What is X's problem?
3. A problem that triggers an internal response: How does X feel?

4. A goal that motivates the character's actions during the narrative: What does X want?
5. Attempts to solve the problems as well as their consequences: What could X do? What do you think will happen next?
6. A resolution to the story: Did X get what he wanted?

They summarized 16 peer-reviewed journal articles published from 1985-2012 looking at inferential comprehension in stories and produced a milestones chart of development in this area: At 3 years children are able to understand internal responses, for example they understand the link between situations and emotions; at 4 years children understand the purpose or goal of the characters and some problems experienced by them; at 5-6 children understand consequences of actions and simple solutions to problems; at 6 years children are able to predict what will come next. However, they concluded that more studies were needed for all of the age groups apart from the 3 year-olds.

Freed and Cain (2016) found a developmental progression in verbal inferential comprehension in a story format in year 3 (aged 7 to 8) and year 5 (aged 9 to 10) children. All of the children performed better when the questions were asked after segments of the story rather than at the end of the complete story. The older children were better at answering questions that involved global and local coherence inferences than the younger children.

1.3.3 Developmental stages of reading inferential comprehension in typically developing children

Table 1.2 Summary of the literature reviewed for reading inferential comprehension in typically-developing (TD) children

Researcher and date	**Inferential comprehension in reading - key findings**
Oakhill (1984)	7-8 year-old advanced readers were better than poorer readers at making inferences from stories.
Cain and Oakhill (1999) and Oakhill and Cain (2012)	Good verbal inferencing skills led to good reading comprehension. Lack of

	general knowledge, poor vocabulary and cognitive delays did not cause inferencing difficulties.
Hudson and Slackman (1990)	Younger children made contextual inferences better than text-based inferences.

Botting and Adams (2005) state that inferencing ability is closely related to educational progress. Verbal inferential comprehension is crucially important for later reading comprehension. Text comprehension is more difficult than verbal comprehension as the decoding of written language places additional cognitive demands on the processing system (Lucas and Norbury, 2015). There is a wealth of literature on the development of inferential comprehension in reading. Norbury and Bishop (2002) point out that research into inferential comprehension in children with reading comprehension has developed in parallel with the research into verbal inferential comprehension. These children are typically developing and are identified by screening of mainstream school populations. The assessment is usually a type of dialogic reading, where the adult asks the child questions whilst reading a classical narrative-structured story (Filiatrault-Veilleux et al, 2015). Many children identified as poor reading comprehenders have age appropriate word-reading and decoding skills, but are significantly impaired in reading comprehension compared with their peers. Verbal and written language are closely inter-related but there is disagreement amongst researchers as to which is the primary cause of difficulty in impaired reading comprehension.

'The Simple View of Reading' (Gough and Tunmer, 1986) states that both decoding skill and verbal language abilities such as vocabulary and grammar predict reading comprehension skills for typically developing children.

Oakhill (1984) reported that 7-8 year-old advanced readers were very good at making inferences. Conversely, children with poor reading comprehension skills had difficulty making inferences from a story they had heard or read. They still had difficulty when the text was available to them during the questions, indicating that their difficulty with inferences could not be attributed simply to working memory difficulties. She suggested that the poor comprehenders' inferential comprehension difficulties reduced their ability to form an integrated and coherent mental model of what they had read or heard.

Cain and Oakhill (1999) tested three groups of children (matched-age skilled and less skilled reading comprehenders, and a verbal comprehension-age matched group) on text-connecting and gap filling inferences. The less skilled reading comprehenders were impaired in making both types of inferences when answering questions from memory compared with peers matched for age and reading accuracy, and significantly worse than younger children matched for reading comprehension ability on text-connecting inferences. When the text was available (so not involving verbal memory), they continued to have difficulty with gap-filling inferences. The results indicated that good verbal inferencing skills were a plausible cause of good reading comprehension ability rather than vice versa. In addition, failure to make inferences was not attributable to lack of relevant general knowledge.

Oakhill and Cain (2012) demonstrated that inferential comprehension predicted reading comprehension ability even after word reading, vocabulary and cognitive ability were taken into account.

Hudson and Slackman (1990, in Adams et al, 2001) discovered that younger children made contextual inferences (requiring integration of existing world knowledge into the text) better than text-based inferences (only recoverable from information in the text) but that text-based inferencing improves with age.

1.3.4 Verbal inferential comprehension difficulties in children with communication impairments.

It has been recognised for over two decades that children with communication impairments have difficulty in making verbal inferences (Adams et al, 2009). Some studies link inferential comprehension difficulties to impaired grammatical understanding, receptive vocabulary and working memory. These are key difficulties for children with specific language impairment, so it is not surprising that these children have inferential comprehension problems (Lucas and Norbury, 2015).

Table 1.3 Summary of the literature reviewed for inferential comprehension in children with communication-impairments (CwCI)

Researcher and date	**Inferential comprehension in CwCI - key findings**
Ellis-Weismer (1985) and Crais and Chapman (1987)	Children with SLI performed like younger language-matched controls on literal and inferential questions from stories.
Bishop and Adams (1992)	There was a trend for 8-12 year-old children with SLI to have more difficulty with inferencing than literal questions.
Leinonen and Letts (1997)	A case study of a girl with PLI from 9:08 to 10.03 showed she had particular difficulty answering inferential questions.
Letts and Leinonen (2001)	Found a developmental progression in inferential comprehension (using questions from the LDA 'What would you do?' cards). The children with LI (mean age 8:07) gave more pragmatically irrelevant and problematic justifications of their answers than the TD controls.
Norbury and Bishop (2002)	Tested 6-10 year-old children on literal and inferential comprehension of a story. 11% of TD; 17% of PLI; 25% of SLI; 70% of HFA children demonstrated poor inferencing skills.
Botting and Adams (2005)	Tested 11 year-olds with SLI and PLI. Found inferential comprehension was in line with general language ability. The PLI group performed worse than the SLI

	group.
Ryder et al (2008)	Children aged 5-11. PLI group performed worse than the SLI group on implicature questions from pictures and stories.
Dodwell and Bavin (2008)	6 year-old children with SLI had difficulty making inferences from story scenarios (linked to their reduced verbal working memory).
Adams et al (2009)	Tested 6-11 year-old children on the ACE inferential comprehension sub-test. Children with SLI and PLI scored the same as comprehension age-matched TD controls. The PLI group performed worse than the SLI group.
Karasinski and Weismer (2010)	Distant inferences were harder for 13-14 year-old language impaired children to understand than controls.
Saladana and Frith (2007)	12-19 year-olds with ASD and reading comprehension difficulties were able to make implicit inferences and draw on relevant world knowledge in a computer 'read aloud' vignettes test.
Norbury and Nation (2011)	Children with ASD with additional LI had more difficulty than children with ASD with no LI and TD controls on a reading inferential comprehension task.
Lucas and Norbury (2015)	Tested 7-12 year-old children's inferential comprehension in a reading test. 12.5% of TD, 33% ASD, over 50% of LI children had difficulty.

1.3.4 Inferencing difficulties in children with Specific Language Impairment (SLI) and Pragmatic Language Impairment (PLI)

Ellis-Weismer (1985, in Bishop and Adams, 1992) compared twelve children with Specific Language Impairment (SLI) to two control groups (one chronological age matched and one language age matched). The children were presented with short story sequences, either as a series of pictures or verbally. They were then asked some Yes/No questions, half of which could only be answered by the child making inferences. The children with SLI were impaired relative to the chronological age-matched controls, and their performance on all conditions resembled that of the younger language age matched controls.

Crais and Chapman (1987) found that language impaired children performed worse than age matched controls, but the same as younger receptive vocabulary matched controls when answering literal and inferential questions about verbally presented short stories.

Bishop and Adams (1992) point out that these two studies indicate that children with language impairments are impaired in general comprehension, literal as well as inferential. Their 1992 study investigated this area further. They assessed sixty one, 8-12 year-old children with SLI and 50 typically developing control children, ten (5 boys and 5 girls) in each of the following age groups: 5, 6, 8, 10 and 12. The children were asked literal and inferential questions about a story they heard or saw as a picture sequence. An example scenario was: 'John was at the beach. He trod on some glass. He had to go to hospital' The inferential questions probing this are: 'Was John wearing shoes?' and 'Why did John go to hospital? They found that the children with SLI were impaired on both tasks (so they were equally impaired when they had the picture supports), even compared to matched comprehension aged children (as assessed on the TROG-1 and the WISC-R verbal comprehension sub-test). However, they did not have more difficulty with the inferential questions than the literal questions (there was a trend for this but it was not statistically significant). It could be a significant factor that 54% of the inferencing questions in this study required the child to refer to a character's mental or emotional state (ie requiring a well-developed theory of mind).

Leinonen and Letts (1997) carried out one of the early studies in inferential comprehension. They documented the communication difficulties of a girl with pragmatic language impairment between the ages of 9:08 to 10:03 and compared her performance on inferential comprehension tasks with two control groups of 8 six year-old and 8 eight year-old typically developing children, 4 boys and 4 girls in each group. The children answered a range of descriptive and inferential questions on 2 composite pictures (‘the park’ and ‘the flood’) and on short stories about ice skating (with pictures) and a piggy bank (without pictures). The results demonstrated that the child with pragmatic impairment (Sarah) had difficulty answering questions that required her to go beyond the picture materials or verbally stated information. For example, she had difficulty suggesting what could have happened before or what could happen next in scenarios. She had much more difficulty answering inferential questions than descriptive questions and she had more difficulty answering the inferential questions than 6 and 8 year-old typically developing control children. This suggested a specific difficulty with inferential comprehension for Sarah. However, the researchers did admit that it was not possible to tell from this study if her main difficulty was with inferencing per se, or if it could be due to working memory, processing contextual information or lack of world knowledge deficits. The current researcher suggests that a significant contributory factor could be that Sarah’s understanding of syntactic structures (as assessed on Bishop’s (1989) Test for Reception of Grammar) was significantly delayed at 8:10 and 10:02 (she obtained an Age Equivalent of only 5:03 on both occasions).

Letts and Leinonen (2001) conducted a further study into inferential comprehension in language-impaired and typically developing children. They tested 14 language-impaired (LI) children aged 7:01-10:03 (mean 8:07) recruited from a specialist school for children with LI in the South of England. All of the children had a language comprehension age of at least 5:06 years. There were ten boys and four girls, in keeping with the sex ratio in LI. Seven of the fourteen children were described as having a language impairment (LI) and seven as having a pragmatic impairment (PI). There were also 3 control groups of typically-developing children (equal numbers of boys and girls): Sixteen 6 year-olds (roughly comprehension-age matched), sixteen 8 year-olds (roughly chronological age matched) and twelve 16 to 17 year-olds, all attending mainstream schools. The latter group were chosen to give a measure of how most adults would answer the task questions. The subjects were asked a range of descriptive (Bishop and Adams’, 1992 ‘literal’) and inferential questions, matched for grammatical complexity, linked to pictures taken from the Learning Development Aids (LDA) 'What would you do?' series (LDA, 1977). Additional questions probed how people were feeling or required problem-solving strategies. The participants were asked to justify the responses they gave. Examples of some of the inferential questions are given below:

Picture 1 – Lost key

A boy is standing outside his house going through his pockets. The door is shut and the assumption is that he is locked out.

Questions:

Does the boy live here? (Inferential)

Is anyone at home? (Inferential)

How does he feel? (Feeling)

What do you think he should do? (Problem-solving)

The study results found a developmental progression in inferential comprehension, with the children with LI performing like the youngest typically developing children. The children with LI also gave more pragmatically irrelevant responses and more problematic justifications of their answers than the control groups. No significant difference in correct number of answers or numbers of problematic justifications was found between the LI and the PI clinical groups.

Norbury and Bishop (2002) compared 6-10 year-old children with specific language impairment (SLI), pragmatic language impairment (PLI), high-functioning autism (HFA) and typically developing children on their inferential processing and story recall abilities. The children with SLI and PLI were recruited from specialist language schools and language units. The stories were taken from Cain and Oakhill's (1999) study but were read out to the children verbally rather than presented in written text. Questions after the story were both literal and inferential. An example of one of the stories ('The Beach Story') and the accompanying inferential questions (taken from the initial Cain and Oakhill (1999) study) is given below:

Debbie was going out for the afternoon with her friend Michael. By the time they got there they were very thirsty. Michael got some drink out of his bag and they shared that. The orange juice was very refreshing.

Debbie put on her swimming costume but the water was too cold to paddle in, so they made sandcastles instead. They played all afternoon and didn't notice how late *it was. Then Debbie spotted the clock on the pier. If she was late for dinner her parents would be angry.*

They quickly packed up their things. Debbie changed and wrapped her swimming costume in her towel. She put the bundle in her rucksack. Then they set off for home, pedalling as fast as they could. Debbie was very tired when she got home, but she was just in time for dinner.

<u>Gap-filling inferences questions</u>

Where did Debbie and Michael spend the afternoon?

How did Debbie and Michael travel home?

The results demonstrated that the three clinical groups had more difficulty answering both literal and inferential questions than their age-matched peers. Initial analysis found that at a group level the children with the communication impairments did not differ to a significant degree in terms of their literal and inferential understanding in a story comprehension task. So, as with Bishop and Adams' (1992) study there were no clear differences amongst the clinical groups in their inferential comprehension abilities. However, further analysis showed that some children did poorly with all question types, whereas some were much worse at answering the inferential than the literal questions. So they computed 'inference deficit' scores for the children by comparing their literal and inferential comprehension. A 'poor inferencing score' was labelled as being below 0.86. Using this new calculation 11% of the typically developing children; 17% of the children with PLI; 25% of the children with SLI; and 70% of the children with HFA demonstrated poor inferencing skills. These revised results indicate that impaired inferential comprehension is not just a consequence of low general ability or weak structural language skills. The current researcher would agree with the finding that the HFA sub-group would have the most difficulties with making inferences but she would have predicted (contrary to Norbury and Bishop's (2002) findings) that the SLI group would perform better than the PLI group. This could be an indication that the SLI group had very poor receptive language skills overall. Only 15% of the inferencing questions in this study required the child to refer to a character's mental or emotional state (39% less than Bishop and Adams' 1992 study). Norbury and Bishop's (2002) conclusion from this study was that children with different communication impairments may represent a continuum of ability and deficit and differ from one another in severity of impairment rather than qualitatively.

Botting and Adams (2005) investigated the semantic and inferencing skills of 25 eleven year-old children with specific language impairment (SLI) and 22 eleven year-old children with Pragmatic language Impairment (PLI). They compared their performance to each other and to three control groups of typically developing children: 35 eleven year-olds, 40 nine year-olds and 37 seven year-olds. The inferential comprehension task was a story comprehension based on the popular children's book 'The Lighthouse Keeper's Lunch' (Armitage, 1994). The children were told a modified story with added inferencing scenarios whilst looking at the pictures. They were then asked a series of Yes/No response literal and inferential questions whilst the pictures remained available to them. This type of response minimises the expressive language demands on the children. Questions could be repeated once if necessary and the children could respond verbally or by gesture. An example of one of the picture texts and 2 related inferencing questions are given below.

Picture 1.

Once there was a lighthouse keeper called Mr Grinling. At night time he lived in a small white cottage perched high on the cliffs. In the day time he rowed out to his lighthouse on the rocks to clean and polish the light.

Inference questions (the children were asked to state if the statements were true or false).

Mr Grinling spent every day in the lighthouse (true)

The light was dirty (true).

The answers were converted into scores for each child. All of the children were also assessed on the Wechsler Intelligence Scale for Children (WISC-IIIR) short form (Wechsler, 1992), the British Picture Vocabulary Scale (BPVS-II), (Dunn et al, 1998), the Test for Reception of Grammar-1 (Bishop, 1989) and the Children's Communication Checklist (Bishop, 2003). However, not all of the participants completed all of these assessments.

The results demonstrated that both groups of children with communication impairments performed significantly below the 11-year-old control group, but not below the 9- and 7-year-old language-similar control groups, suggesting that inferential comprehension was in line with general language ability. Six children in the PLI group who met diagnosis for autism performed more poorly than the other two clinical groups on both tasks, but not statistically significantly so.

Ryder et al (2008) use 'implicature' instead of inferencing. The term is based on Sperber and Wilson's (1997) Relevance Theory and is defined as the ability to use reasoning to integrate contextual information to work out intended meaning. They carried out a study investigating implicature skills in 99 children: 18 with SLI , 9 with PLI and two groups of typically developing children (32 children aged 5-6 years and 40 children aged 7-11 years). They tested language comprehension using questions of increasing pragmatic complexity in different verbal contexts (scenarios with and without pictures and a story with supporting pictures). The most pragmatically demanding questions involved making inferences.

An example of a picture-based implicature question, where three children are playing outside and one of them is playing in the road, is: Maxine was playing outside. Maxine's mum was worried Maxine might get hurt. Can you point to Maxine?

An example of a verbal only implicature question is: Marie was going to her piano exam. Marie felt sick. Why did Marie feel sick?

The results demonstrated that both the children with SLI and the younger children were not able to accurately answer the inference questions. In addition, the children with PLI performed significantly worse than the SLI children on these questions and they had a particular difficulty with integrating information from the context. However, the sample sizes of the SLI and particularly the PLI group are much smaller than the control group sample sizes, indicative of the small numbers of children with SLI and PLI in the general population and hence the difficulties recruiting large numbers of these children.

Dodwell and Bavin (2008) used 'The Birthday Story' (Culatta et al, 1983) and 'The Fish Story' from the ERNNI narrative assessment (Bishop. 2004) to test 6 year-old children with SLI's narrative, memory and inferencing skills. They found that the SLI children had difficulty making inferences from the story scenarios. They also reported that their impaired narrative abilities were linked to their reduced verbal working memory.

Adams et al (2009) investigated inference and sentence comprehension in sixty-four 6 to 11 year-old children with language impairments (CwLI), thirty-eight of whom had specific language impairment (SLI) and twenty-six had pragmatic language impairment (PLI), as identified by the children's speech and language therapists. They also recruited two control groups of sixty-four age-matched and sixty-four sentence comprehension age-matched typically developing children. The children were required to complete a sentence comprehension test and an inferential comprehension test (both of which were preliminary versions of the corresponding sub-tests for the ACE 6-11 language assessment, Adams et al, 2001) (see the current assessments summary in section 7 for a detailed description of the kitchen burglary inferential comprehension sub-test). Both experimental groups obtained lower scores on the inferential comprehension task than the age-matched control group but similar scores to the sentence comprehension age-matched control group. This suggests that the representations of events and characters of the children with communication impairments were similar in nature to children with the same comprehension age. The children with PLI scored less on the inferential comprehension task than the children with SLI and significantly less than their sentence comprehension-matched controls. They also performed more poorly on the developmentally more complex inferential comprehension questions. Adams et al (2009) concluded that children with PLI are more likely than children with SLI to have difficulty with a story-plus-question-type inference comprehension task. However, there was overlap in performance between the SLI, PLI and comprehension-matched control groups and the study was unable to identify particular sub-group specific inferential comprehension error patterns. The authors acknowledge that they did not systematically measure world knowledge, vocabulary, social understanding, social flexibility or working memory in this study.

Karasinski and Weismer (2010) tested 4 groups of 8th grade (aged 13-14 years) pupils on comprehension of adjacent and distant inferences in a story context. Adjacent inferences occur immediately after given information in the text. Distant inferences refer to information further back in the text and therefore depend on a good working memory. 527 children in total were tested. The four groups were: normal language (NL, scoring within the average range on both nonverbal and verbal tests), low cognition (LC, scoring below 87 on a nonverbal assessment but in the normal range on the language tests), specific language impairment (SLI, scoring in the normal range on a nonverbal assessment but in the impaired range on the language assessments) and non-specific language impairment (NLI, scoring in the impaired range on both language and nonverbal tests). The results demonstrated that distant inferences were significantly more difficult to understand for all 4 groups, but the NL group outperformed the other 3 groups on these. The LC group performed better than the NLI group. The authors interpreted these results as evidence that language comprehension, verbal working memory and general world knowledge are essential for the ability to make distant inferences. The current researcher is not convinced about the validity of the LC and NLI diagnostic groups as they are both children with cognitive delays, the only difference being that the LC group scores better on the language tests than the NLI group

1.3.6 Inferencing difficulties in children with Autism Spectrum Disorder (ASD)

Some researchers suggest that inferencing difficulties are linked to a theory of mind deficit. This is a key feature of children with ASD, so, if this is the case, it is not surprising that they have difficulties with inferential comprehension, in particular those inferences that require understanding someone else's thoughts and behaviours. Approximately a third of children with ASD have difficulties with reading comprehension and they do worse on texts that include social demands. However, as for LI and PLI, overall language competence is still the greatest predictor of inferential comprehension ability in ASD (Lucas and Norbury, 2015).

1.3.7 Reading inferential comprehension difficulties in adolescents with ASD

Saladana and Frith (2007) carried out an experiment with 16 male adolescents with ASD, aged 12:04-19:04 (mean age of 14:09), who had good reading accuracy but poor reading comprehension. They tested whether this group were able to make bridging inferences from their world knowledge whilst reading. They included a control group of typically developing adolescents aged 11:06-18:08. The tests were carried out on a computer and they measured real-time 'on-line' inferencing ability. Participants read aloud two-sentence vignettes which included either physical content or social content inferences. The inference questions were primed or non-primed. Both groups read the inference-primed questions more quickly than the non-primed questions, and no difference was found between the physical content and social content inferences. These results indicate that, for this cohort of adolescents, the reading comprehension difficulties in ASD cannot be attributed to an inability to make implicit inferences or draw on relevant world knowledge. However, Lucas and Norbury (2015) point out that the participants with ASD varied hugely in their receptive vocabulary ability, with standard scores ranging from 53-147. Consequently the adolescents with ASD with the higher receptive vocabulary scores could have masked the difficulties of those with the lower receptive vocabulary scores by significantly increasing the mean score of the inferential comprehension task.

Norbury and Nation (2011) compared the performance of children with ASD and no structural language difficulties (ALN) with children with ASD with language impairments (ALI) and a control group of typically developing children (TD) on a text comprehension task (asking both literal and inferential questions). Both ASD groups correctly answered a similar number of literal questions to the TD group However, the ALI group had more difficulty than both the other groups in answering the inferential questions. These results indicate that it is overall language ability that is key for good inferential comprehension. More specifically, the authors reported that verbal comprehension predicted 30% of inferencing competence with ASD status predicting only 10% of additional variance.

1.3.8 Reading inferential comprehension in children with Autism Spectrum Disorder (ASD) and Language Impairment (LI)

Lucas and Norbury (2015) tested eighty-six 7 to 12 year-old children with the Neale Analysis of Reading Ability (Revised) (Neale, 1997) and compared their responses. They used Tager-Flusberg and Joseph's (2003, in Lucas and Norbury, 2015) ASD phenotypes: ALI (children with ASD with LI) and ALN (children with ASD and no language difficulties). The final participants were: 12 children with language impairment (LI), 27 with autism spectrum disorder (ASD) and no structural language difficulties (ALN), 15 children with co-occurring ASD and language impairment (ALI), and 32 typically developing (TD) children. They found that the inferential comprehension questions in the test were difficult to answer for 12.5% of the TD children, 33% of the children with ASD, and over 50% of the children with LI. The strongest predictor of reading inferential comprehension ability was verbal language skills, followed by vocabulary knowledge and verbal working memory. Children in all three clinical groups had more specific deficits in inferencing than the TD group However, Lucas and Norbury (2015) describe inferential understanding as a pragmatic language skill and they therefore exclude it from their definition of LI. Many other researchers would disagree with this definition and would include inferential comprehension within receptive language and therefore within the LI umbrella. In addition, their only measures for LI were receptive and expressive vocabulary and the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals (Semel et al, 2006), which tests verbal auditory memory more than expressive language. Consequently, their findings are very difficult to interpret.

1.3.9 Critical evaluation of the above studies

Current findings on inferencing abilities in children with developmental disorders are contradictory. Some studies report that children with communication impairments have difficulties with both literal and inferential comprehension but some studies indicate a specific difficulty with inferential comprehension. Working memory, vocabulary and world knowledge deficits have been demonstrated to affect inferential comprehension ability in some studies but not others.

Leinonen and Letts (1997) highlight that children with language and communication difficulties may perform differently in different contexts or at different times due to the nature of the task, the context and the effect of interactive partners. Participant characteristics, the complexity of the inferences and methodological issues can also affect inferential comprehension ability.

Participant characteristics

Lucas and Norbury (2015) suggest that some conflicting findings could be due to different participant characteristics. Botting and Adams (2005) state that the lack of clearly defined subgroups of children with communication disorders has made it difficult to evaluate and implement research findings. They report that their failure to find any clinically significant difference in inferential comprehension between children with PLI and HFA could be in part due to the heterogeneity of both groups and the ever-widening definition of ASD. Adams et al (2009) acknowledge that they relied on practitioner diagnosis of the social communication status of the children in their study as there are no clear cut criteria to differentiate SLI from PLI and the boundaries of PLI merge into other diagnostic groups. They recommend that future studies in this area supplement clinicians' diagnoses with a less subjective diagnostic tool such as the Children's Communication Checklist-2 (Bishop, 2003). The children's processing and cognitive profile (working memory, vocabulary and world knowledge) could also influence their ability to make inferences. Bishop (1997) suggests that the causes and types of inferencing difficulties are likely to vary from child to child.

<u>Complexity of inferences</u>

Leinonen and Letts (1997) point out that inferential questions can have different levels of complexity depending on the number of inferences to be drawn, the familiarity and obviousness of given information, and whether supporting visual material is provided. Norbury and Bishop (2002) state that the type of inference being assessed can make a difference (eg inferences that involve theory of mind are more difficult for certain sub-groups of children than inferences that do not).

<u>Methodological problems</u>

Norbury and Bishop (2002) propose that some of these different findings may have more to do with the methodology used for testing inferencing. Letts and Leinonen (2001) highlight that one of the problems with testing inference is that it is an internal process assessed by a child's external response. Also, scoring of acceptable responses can be subjective, with seemingly illogical responses being generated from a logical internal thought sequence from the child's perspective. Adams et al (2009) suggests that children with PLI could still have specific difficulty with verbal inference but that the experimental paradigms deployed up to now have failed to detect these. They also point out that the standardised tests used may not be sensitive enough to measure the children's specific comprehension difficulties. In their pre-school scoping review Filiatrault-Veilleux et al (2015) found that no two studies used the same methods to assess inferential comprehension, thus making it extremely difficult to record developmental progress in this area with any accuracy.

All of these features demonstrate how complex and difficult it is to control for all possible variables in studies into inferential comprehension. It is hoped that the development of a standardised British assessment of inferential and idiom comprehension will help to control some of these variables (eg the complexity of the inferences and some of the methodological difficulties).

This section has examined inferential comprehension abilities and difficulties in typically developing children and children with communication impairments. The next section summarises the literature on idiom comprehension abilities and difficulties in typically developing children and children with communication impairments.

CHAPTER ONE SECTION FOUR: 1.4 Idiom comprehension in typically developing children and children with communication impairments

This section begins with a chart summarising the research studies in each area. Each study is then described in detail below the corresponding chart. The studies are critically evaluated together in section 1.4.3.

Table 1.4 Summary of the literature reviewed for idiom comprehension in typically-developing (TD) children.

Researcher and date	**Idiom comprehension in TD children - key findings**
Levorato and Cacciari (1995)	The acquisition of figurative language takes place between 7-11 years.
Gibbs (1991) and Levorato and Cacciari (1999)	Older children (8-9 years) understand transparent idioms out of context better than younger (6-7 years) children.
Caillies and Le Sourn-Bissaoui (2008)	French 5-6 year olds can understand transparent idioms in context. 7-8 year-olds understand opaque idioms. Opaque idiom comprehension was predictable from administered theory of mind tasks.
Cain et al (2009)	5 year-old British children can use semantic analysis to understand transparent idioms in context. Idiom knowledge is still not fully developed by 11-12 years.

1.4.1 Typical development of idiom comprehension

Whyte et al (2014) and Norbury (2004) state that idioms are an important area of research as they occur frequently in verbal and written language and are used extensively in the language of the classroom. Good comprehension of idioms correlates with good academic ability. There are two main theories of idiom acquisition in typically developing children: Ackerman's (1982, in Norbury, 2004) 'giant lexical unit' theory which suggests that idioms are acquired in the same way as single lexical items and Gibbs' (1987, in Norbury, 2004) 'metasemantic theory' which suggests that the meaning of some idioms can be gleaned from analysing their component parts. If participants don't rely on context to understand idioms, this suggests that they have been stored as a giant lexical unit. Older children and adults do not rely on context to understand idioms as much as younger children do. Both theories have validity at different stages of development and for different types of idioms. Conner et al's (2011) study of idiom comprehension in adulthood (age bands 18-30 and 60-85) supports the 'giant lexical unit' theory as the idiom retrieval difficulties in the older adults correlated with their lexical item retrieval difficulty. Idiom acquisition involves a combination of top-down processing (from the context) and bottom-up processing (semantic analysis). Frequency of exposure is also a significant factor in idiom comprehension. Highly familiar idioms are easier to understand then less familiar idioms. In addition, transparent idioms are easier to understand than opaque idioms.

Children use both semantic analysis and supporting context to process the meaning of idioms. There are conflicting findings about the age at which children can use semantic analysis to understand idioms. Several studies indicate that this is an early developing skill, but others claim it develops later in childhood. Levorato and Cacciari (1995, in Cain et al 2009) propose that the most significant development of the skills and abilities needed to process and acquire figurative language takes place between 7 and 11 years of age. Their 'global elaboration model' illustrates that the understanding of idioms depends on the same skills and strategies that underpin general verbal and reading comprehension.

Gibbs (1987, 1991 in Caillies and Le Sourn-Bissaoui, 2008) demonstrated that that transparent idioms (which have literal components, such as 'lay down the law') are easier to understand than opaque idioms (where the meaning cannot be derived from the individual words, such as 'kick the bucket') due to the closer link between figurative and literal meaning in the former. Consequently, children understand transparent idioms earlier than opaque ones. Nippold and Taylor (2002) point out that familiarity of idioms is also a factor in their comprehension.

Developmental studies have shown that idioms that are higher in familiarity and transparency are easier to understand than idioms that are less familiar and more opaque.

Specific critical evaluation is given after the description of some studies and a generic critical evaluation summarising features of many of the studies is given at the end of this section.

Gibbs (1991, in Cain et al, 2009) demonstrated that third-graders (mean age 8:09) but not first graders (mean age 6:10) were able to understand transparent idioms out of context. Levorato and Cacciari (1999, in Cain et al, 2009) found the same results with 9 and 7 year-old children. Gibbs also found that all age groups benefited from the presence of context. Only 8 and 9 year-olds were able to use semantic analysis to aid comprehension of some transparent idioms out of context. The 8 year-olds understood 37% and the 9 year-olds 42% of the idioms tested.

Caillies and Le Sourn-Bissaoui (2008) reported that children as early as third-kindergarten grade (aged 5-6 years) were able to understand transparent idioms in context in a multiple choice task, while they needed to be in second-grade (aged 7-8 years) to understand opaque idioms.

However Nippold and Rudzinski (1993, in Cain et al, 2009) and Nippold and Taylor (1995, in Cain et al, 2009) found a positive correlation between transparency and performance on an idiom explanation task for 14 and 17 year-olds but not for 11 year-olds, indicating that semantic analysis of idioms develops quite early but continues to be refined into later childhood.

Caillies and Le Sourn-Bissaoui (2008) tested twenty-six 5 year-old (mean age 5:03), thirty 6 year-old (mean age 6:03) and 25 seven year-old (mean age 7:04) children's theory of mind level and understanding of transparent and opaque French idioms. The children had to complete five theory of mind tasks: an appearance–reality task, three false-belief tasks and a second-order false-belief task. They then had to perform a multiple choice task after listening to the idioms in context. The results demonstrated that only the opaque idiom comprehension was predicted from the theory of mind scores (particularly from the second-order false-belief task). The disadvantage of this for the current assessment is that the idioms tested were French and they cannot be directly translated for use in an English assessment.

Cain et al (2009) tested transparent and opaque idiom comprehension in 40 British children (23 girls and 17 boys): 20 year 3 children (mean age 7:10), 20 year 5 children (mean age 9:11); and 19 adults (12 women and 7 men, mean age 19:04) from the North West of England. 24 idioms were used in the test. 12 were common British idioms and 12 were novel idioms that were translations of real non-English European idioms for which there are no equivalents in British English. 6 were familiar transparent idioms (eg 'to skate on thin ice'); 6 were familiar opaque (eg 'to take the biscuit'); 6 were novel transparent (eg 'to try to make a hole in water') and 6 were novel opaque (eg 'to whistle in your thumb'). Idioms were presented in and out of a supportive story context and were assessed via a multiple-choice test. They found that even the 5 year-olds were able to use semantic analysis to understand the transparent idioms and were sensitive to meaning in context. They reported that children age 7 years and under benefited from transparency only when idioms are presented in context, whereas 9-year-olds benefited from transparency when idioms were presented out of context as well. However, they also demonstrated that idiom knowledge was still not fully developed in 11 and 12 year-olds. The use of non-British idioms in this study means that the results cannot be compared to the current researcher's study which uses only familiar British idioms.

<u>Factors influencing idiom comprehension</u>

The choice of tasks used to measure idiom comprehension may influence the results. Idiom explanation tasks may disadvantage younger children due to their less well developed expressive language and comprehension skills compared to older children (Cain et al, 2009). The preferred method of assessing idiom comprehension is therefore by using a multiple-choice task to test idioms in context.

1.4.2 Studies of idiom comprehension in children with communication impairments

Table 1.5 Summary of the literature reviewed for idiom comprehension in children with communication impairments (CwCI)

Researcher and date	**Idiom comprehension in CwCI - key findings**
Kerbel and Grunwell (1998)	'Acting out the idiom' task. Children

	with PLI had more difficulty than TD children.
Vance and Wells (1994)	LI group performed the same as language-matched controls on idiom and metaphor comprehension.
Rinaldi (2000)	11-14 year old children with LI performed worse than language-matched children on ambiguous language comprehension (including idioms).
Dennis et al (2001)	8 children with HFA failed to understand metaphor and speech acts on the Test of Language Competence.
Whyte et al (2014)	5-12 year-old children with ASD performed as well as syntax-matched controls on an idiom definition task. Advanced TOM ability was related to idiom comprehension in the children with ASD only.
Qualls et al (2004)	13-15 year-olds with LI performed significantly worse than TD peers on an idiom comprehension task.
Lee et al (2015)	6-11 year-old Korean children with HFA and ADHD performed significantly worse than TD controls on idiom comprehension.

Cain et al (2009) highlight that comprehension of figurative language, such as idioms, causes particular difficulties for children with language impairments. Kerbel and Grunwell (1998, in Norbury, 2004) tested idiom comprehension in children with pragmatic language impairment (PLI) using a definition task and a prop-based ‘acting out the idiom’ task (to limit the need for expressive language).

The children with PLI demonstrated difficulties compared to the controls in the latter but not the former task. Unfortunately, the language levels of the children in both groups were not given and the PLI group also included some children on the autism spectrum and it is well-documented that these children have difficulty with pretend play, which could account for the difference.

Vance and Wells (1994 in Norbury, 2004) found no deficit in idiom and metaphor comprehension in children with language impairment when they were compared to younger language-aged matched controls. This finding suggests that idiom comprehension develops in line with general verbal language skills. A criticism of this study is that the response mode was the choice of one out of three pictures, one of which was a depiction of the literal meaning of the idiom, making it an easier task as process of elimination could be employed.

Rinaldi (2000), on the other hand, reported that 11-14 year-old children with language impairments were impaired on ambiguous language comprehension (which included idioms) compared to language-matched controls. A criticism of this study is that Rinaldi used a complex response mode which also included pictures of the literal foils.

<u>Idiom comprehension in children with autism spectrum disorder (ASD)</u>

Even though clinicians and researchers often report huge difficulties with non-literal language comprehension in children with ASD, there are actually very few studies that investigate this. Specific difficulties with theory of mind have been extensively identified however, and these contribute to the difficulties understanding implied meaning, idioms, metaphors, irony, lies, jokes, deception and faux pas (Dennis et al, 2001). Whyte et al (2014) outline the two competing explanatory theories for the difficulties understanding figurative language in ASD. One is relevance theory. Happe (1993) proposes that an understanding of speakers' intentions (a key feature of relevance theory) is crucial for understanding figurative language whereas other researchers such as Gernsbacher and Pripas-Kapit (2012) and Norbury (2004) propose that basic language abilities, in particular understanding vocabulary and syntax are the most important prerequisites for idiom comprehension. Gernsbacher and Pripas-Kapit (2012) stated that if children with ASD "don't have difficulty comprehending language in general, they don't have difficulty comprehending

metaphoric language in particular" (p94). They also pointed out that receptive vocabulary was a poor matching variable to control for structural language abilities.
In her advanced theory of mind study, Happe (1994) found that individuals with ASD scored the same as typically developing controls on two stories that contained idioms. She interpreted this as meaning that theory of mind was not necessary to understand idioms as they could be learned as 'frozen expressions'. However, this does not explain how the idioms are initially acquired.
Dennis et al (2001) used the Figurative Language Sub-Test of the Test of Language Competence (Wiig and Secord, 1989) to assess 8 children with high functioning autism (HFA). The results indicated that these children could define words and identify multiple meanings for ambiguous words. However, they failed to understand metaphor and to produce speech acts, both of which are inferences that are crucial for successful social communication because they elaborate meaning or convey intentions. The results also highlighted that it is understanding the speaker's intentionality in social contexts that is most impaired in children with HFA. Criticisms of this study, as summarised by Norbury (2004), are that the verbal IQ of the children ranged hugely (from 71 to 146) and there was a lot of within-group variation on the sub-test scores. Consequently it is not clear from this study if idiom comprehension is linked to general language ability or not.
Whyte et al (2014) carried out a study looking at the links between idiom comprehension, theory of mind and syntactic skills in 26 children with ASD (aged 5-12) and 2 control groups: 1 consisting of 26 chronological age- and nonverbal IQ-matched children and the other consisting of 26 syntax-level matched children. The children were asked to verbally define the meaning of 20 verbally-presented idioms in the context of a short supporting paragraph. Syntactic ability was measured with the Syntax Construction sub-test of the Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999) which is a test of expressive syntax. Advanced theory of mind ability was assessed with six of Happe's mentalising Strange Stories (O'Hare et al, 2009) and the children's version of the Reading the Mind in the Eyes (RMTE) task (Baron-Cohen et al, 2001). The Strange Stories task requires the interpretation of the intentions and mental states of characters in short stories and involves the recognition of lies, persuasion, forgetting, etc. The RMTE requires the children to match complex mental state words (jealous) or short phrases (thinking

about something) to photos of facial expressions from the eye region of different faces. Both tests are appropriate for children aged 5-12 years. The results demonstrated that the children with ASD performed worse than the age-matched controls but as well as the syntax-matched controls on the idiom comprehension task. This is a surprising finding given that children with ASD usually have a strength in the area of expressive syntax. In addition, advanced theory of mind abilities were related to understanding of idioms in children with ASD but not the typically developing children.

<u>Studies of children with mixed communication impairments</u>

Qualls et al (2004) compared the understanding of idioms in 27 eighth grade children (13-14 year-olds) with language-based learning disabilities (LBLD) to that of 22 typically developing peers (mean age 13.8). The results demonstrated that the children with LBLD performed significantly worse than the typically developing control children. Reading ability was associated with comprehension of the low familiarity idioms but only when they were presented in a story context.

Norbury (2004) explored the understanding of idioms in isolation and in context in 93 children with communication impairments and 39 age-matched peers. Due to the overlap between the diagnostic sub-groups of SLI, PLI, ASD and Asperger's syndrome she grouped the children by language ability. The 4 clinical experimental groups were: 29 children with language impairment (LI), 6 children with pragmatic impairment (PI), 29 children with ASD plus LI (ASL) and 29 children with ASD only (ASO). Norbury hypothesises that the theory of mind and central coherence impairments in children with ASD should make it more difficult for them to do well on idiom comprehension tasks that require the processing of context to infer figurative meaning. The results demonstrated that all of the children in the study understood idioms in context better than in isolation. However, the LI and ASL groups did not benefit from context as much as the other groups. This indicates that language ability was a bigger predictor of performance with idioms in context than pure ASD. Age and working memory were other significant predictors. Norbury (2004) only used 10 idioms in this study, 5 transparent and 5 opaque. She deliberately selected low familiarity idioms to test the use of context to understand them. Her scoring system of using a 3 point scale ranging from 0 (don't know or literal answer) to 2 (correct) is open to interpretation. She only gives one example of what a score of 1 would entail for 1 idiom (burning the candle at both ends – 'he's tired').

Lee et al (2015) found that 6-11 year-old Korean children with high-functioning Autism Spectrum Disorder and children with ADHD both performed significantly worse than typically developing children on their Korean idiom comprehension test.

1.4.3 Critical evaluation of the above studies

The above studies present conflicting findings. One reason for this is that it is very difficult to control for all variables in studies of idiom comprehension. These variables include:

- different sub-groups of communication impairment in the participants
- varying numbers of participants
- different cognitive and linguistic levels of the participants
- the methodology used
- the familiarity and transparency of idioms
- the use of context or not
- the mode of response employed

These confounding variables make it very difficult to identify the specific cause of any idiom comprehension difficulties.

Findings from studies on the relationship between theory of mind and figurative language (Happe 1993, Caillies and Le Sourn-Bissaoui, 2008) have been contradicted by findings on the link between general language skills and figurative language (Norbury, 2004, Gernsbacher and Pripas-Kapit, 2012). One flaw in the former theory is that complex Theory of Mind (TOM) is strongly dependent on well-developed verbal language skills and many TOM studies have not matched children according to language level.

The next section examines the inferential and idiom comprehension abilities of children with other types of disorders. It will give further support to the above findings from children with primary communication disorders.

1.4.4 Other disorders: Inferential and Idiom comprehension difficulties

Table 1.6 Summary of the literature reviewed for Inferential and Idiom comprehension in children and adults with other disorders

Researcher and date	**Inferential and Idiom comprehension in other disorders**
McInnes et al (2003)	9-12 year-old boys with ADHD had more difficulty with inferential comprehension than TD peers.
Crespo et al (2007)	6-13 year-old Chilean children with ADHD had significantly more difficulty understanding non-literal meanings – particularly idioms – than TD controls.
Shields (1991)	People with right-hemisphere (RH) lesions had difficulties with inferential comprehension.
Ferstl et al (2002)	People with left frontal lobe impairments demonstrated inferential comprehension difficulties.
Dennis and Barnes (2001)	Children with severe Closed Head Injury (CHI) had more difficulties with inferential comprehension than children with mild CHI or TD controls.
Ferstl et al (2005)	Patients with RH damage were impaired in understanding implicit main ideas. Patients with Traumatic Brain Injury (TBI) were most impaired in understanding implicit information.
Huber-Okrainec et al (2005)	7-18 year-olds with spina bifida meningomyelocele (SBM) had more difficulty understanding opaque idioms than TD children.

Dennis and Barnes (1993)	6-15 year-old children with early onset hydrocephalus (EOH) had difficulty making inferences.

1.4.4.1 Attention Deficit Hyperactivity Disorder (ADHD)

McInnes et al (2003) found that 9-12 year-old boys with ADHD had more difficulty understanding inferences than children without ADHD. They also showed significantly poorer verbal working memory, spatial span, and spatial working memory.

Crespo et al (2007) carried out a study with twenty-nine Chilean 6-13 year-old children with ADHD and a control group of children without ADHD on understanding of literal meanings and non-literal meanings (indirect speech acts and idioms) in cartoons. The children with ADHD had significantly more difficulty identifying the non-literal meanings (particularly idioms) than the control group. In addition, the scores of the children without ADHD on the non-literal comprehension increased with age, whereas the scores of the children with ADHD did not.

1.4.4.2 Brain injury

Closed head injury (CHI)

Shields (1991) reports that the communication profile of adults with right hemisphere lesions is similar to that of children with semantic-pragmatic difficulties. Both groups use fluent and grammatically complex language, but do not pick up all the cues from communicative situations, and both fail to make use of paralinguistic features or to comprehend inferential meaning. Ferstl et al (2002) also report difficulties with inferential comprehension in patients with left-frontal lobe impairments.

Dennis and Barnes (2001) assessed three forms of text comprehension: literal understanding literal information), inferential understanding (making pragmatic inferences, textual coherence inferences, or enriching inferences), and understanding people’s mental states and intentions (producing speech acts, appreciating irony, and understanding deception) in children with mild or severe closed head injury (CHI). The children with severe CHI were significantly impaired on tasks of literal text understanding, inferencing, and intentionality. The children with mild CHI understood the literal questions but were impaired on some of the inferencing and all of the intentionality questions. Barnes and Dennis (2001) also found that the 18 children with severe CHI in their study had significantly more difficulties with inferential comprehension than the 15 children with mild CHI or the 18 age-matched controls.

Traumatic Brain Injury (TBI)

Moran et al (2006) tested ten young people, aged 12 to 21 years who had suffered a traumatic brain injury (TBI) prior to the age of 10 years, and 10 typically developing individually age-matched peers on working memory and comprehension of low-familiarity proverbs. They found that the adolescents with TBI performed significantly worse than the controls on the proverb comprehension but that working memory capacity influenced performance in all of the participants.

CHI and TBI

Ferstl et al (2005) assessed story comprehension in 96 brain damaged patients and compared the results with 49 typical controls. The whole group of patients made more errors overall but had particular difficulty with questions involving inferential comprehension. The 18 patients with left-hemisphere damage had particular difficulty with stated details, the 12 patients with right-hemisphere damage had most difficulty understanding implicit main ideas, and the 34 patients with traumatic brain injury (TBI) were most impaired in understanding implicit information. Performance on understanding implicit main ideas correlated with neuropsychological tests of executive functions. The performance on the other three question types correlated with long-term memory and verbal ability.

Huber-Okrainec et al (2005) studied idiom comprehension in children with spina bifida meningomyelocele (SBM), a neurodevelopmental disorder associated with agenesis and hypoplasia of the corpus callosum. Participants were first-language English-speaking children: thirty-eight with SBM (age range 7.25-18.67) and thirty-eight age- and gender-matched typically developing control children. The children with SBM had been born with the condition and treated for hydrocephalus with a diversionary shunt shortly after birth. A common feature of children with SBM is problems with discourse comprehension. The results demonstrated that, compared to the control children, the children with SBM had much more difficulty understanding opaque idioms, but they did understand many transparent idioms. As transparent idioms are processed more like literal language the authors interpret this a result of their impaired interhemispheric connections. The current researcher's criticism of this study is that it covers a very wide age range (a span of 11 years), with averaged results. This makes it very difficult to extract information about the performance of specific age groups.

Early Onset Hydrocephalus

Dennis and Barnes (1993) tested 101 children, ages 6 to 15 years (50 with early-onset hydrocephalus (EOH), and 51 typically developing control children), on four oral discourse tasks: establishing alternate meanings for ambiguous sentences, making bridging inferences, and understanding figurative expressions and producing speech acts. Children with hydrocephalus performed more poorly than the controls on all four tasks. Barnes and Dennis (1998) demonstrated that children with EOH have difficulty with discourse skills where context is needed to derive meaning. In a familiar story-retelling task the children with EOH produced less core semantic content than the control children. They also had difficulty making inferences, interpreting novel figurative expressions and recalling factual information from the story.

The next section will examine social inferencing and its links to theory of mind skills in typically developing and communication-impaired children. Some of these findings have been briefly mentioned in the above sections but they will now be expanded upon.

CHAPTER ONE SECTION FIVE: 1.5 Social inferencing in typically developing children and children with communication impairments

This section begins with a chart summarising the research studies in each area. Each study is then described in detail below the corresponding chart. Critical evaluation is given after the description of each study.

Table 1.7 Summary of the literature reviewed for social inferencing in typically-developing (TD) children and children with communication impairments (CwCI)

Researcher and date	**Social inferencing in TD children and CwCI - key findings**
Ford and Milosky (2003)	Children with LI's impaired emotional inferencing correlated with their delayed verbal comprehension.
Happe, (1994) and Jolliffe and Baron-Cohen (1999)	Young TD children and adults with HFA had difficulty with the mentalising questions in the 'Strange Stories Test'
Kaland et al (2005)	10-20 year-old Danish people with Asperger's Syndrome (AS) performed less well on the mental state inferences in the Strange Stories Test than the TD controls.
O'Hare et al (2009)	5-12 year-old children demonstrated a developmental order of understanding in the Strange Stories Test.
Eisele et al (1998)	4-17 year-old children with brain injuries had difficulty understanding mental state verbs.
Dennis et al (2001)	Children with HFA had difficulty understanding and using mental state words.
Spanoudis et al (2007)	Cypriot children with PLI and SLI had difficulty understanding mental state

	verb tasks. The PLI group were more impaired than the SLI group.
Baron-Cohen et al (1999)	Performance of 7-11 year-old TD children on the Faux Pas test improved with age. The children with HFA/AS were more impaired on this task than their TD controls.
Banerjee and Watling (2005)	9 year-old TD children performed better than 6 year-olds on faux pas tasks.
Pearson and Pillow (2016)	Faux pas are usually understood by 7 year-old TD children.
Channon et al (2005)	Children with CHI were impaired on the mentalising stories

1.5.1 Recognition of emotions

Ford and Milosky (2003) tested American kindergarten children (up to 6 years) with language impairment (LI) and age-matched controls (CA) on their ability to correctly identify and label facial expressions depicting four basic emotions (happy, surprised, sad and angry) in isolation and in a story context. All of the children were able to identify and label the facial expressions in isolation, but the children with LI had more difficulty than the CA group integrating emotional knowledge with event context in order to infer a character's feelings. The LI children's impaired emotional inferencing ability correlated with their lower scores on a standardized test of language comprehension. The disadvantage of this study is that the normative findings apply only to American children.

1.5.2 Social inferencing and Theory of Mind (TOM)

Social inferencing is vital for effective social interaction. It is the ability to make inferences in social situations and it requires the well-developed social-cognitive competence of Theory of Mind (TOM). This is the ability to work out what other people are thinking and feeling based on their verbal and nonverbal behaviour and social cues. Baron-Cohen et al (2013) define it as the ability to explain and predict one's own and others' mental states. A TOM is essential for individuals to be able to make predictions about other people's behaviour (Hadwin et al, 2015). Wimmer and Perner (1983) developed the earliest 'false belief' task – a mother hiding her son's chocolates when he was not looking. False-belief understanding is required for a child to correctly predict that the boy will look for his chocolates where he thinks they still are rather than in their new location. In the 1980s, Baron-Cohen and his colleagues developed a similar 'Sally-Anne' false belief task (Baron-Cohen et al, 1985). These first-order TOM tests assess whether the child understands that other people can have different beliefs to them and that their beliefs can sometimes be incorrect.

There is a clear developmental progression in TOM acquisition in typically developing children: the understanding of desires comes before the understanding of false beliefs, the understanding of the absence of knowledge develops after false beliefs, and thinking about thinking develops last (Baron-Cohen et al, 2013). Pearson and Pillow (2016) point out that as understanding of mental states advances, children are able to comprehend increasingly subtle and complex social cues.

Some three year-olds recognise that other people do not see or know the things that they do and that their desires can be different from other people's. Under four years of age, children do not understand that there might be alternative representations of the same object (eg a sponge that looks like a rock).

First-order TOM tests are passed by typically-developing children of 4–5 years of age (O'Hare et al, 2009). Second-order TOM tasks assess 'thinking about thinking'. They involve working out what someone thinks someone else believes in story scenarios. Typically developing children acquire second-order TOM at around 6-7 years of age (Hadwin et al, 2015). From about 6 to 8 years of age, children recognize that two individuals may interpret the same information in different ways (Pearson and Pillow, 2016).

One of the main criticisms of the TOM tests is that they require a relatively high level of verbal comprehension to pass them. For example, in the Sally-Anne test, embedded sentences such as 'Anne thinks that the ball is in the box' need to be understood. Several studies have found links between level of theory of mind and level of language comprehension. However, it is likely that the two skills are inter-dependent (Hadwin et al, 2015). In fact, there is a lot of evidence on the close link between theory of mind and language development in general In addition, theory of mind develops alongside empathy, knowledge of social rules and awareness of self and others (Spanoudis et al, 2007). Leinonen and Letts (1997) propose that social cognition deficits rather than linguistic difficulties cause the pragmatic impairments in children with high functioning ASD (HFA) and pragmatic language impairment (PLI). They suggest that frontal lobe executive functioning ability, which controls goal-directed and purposeful activity and the ability to integrate information into a coherent whole, is a key cognitive skill required for inferencing. These findings link in with the research findings of Ferstl et al (2002), given in section 4 above, on the inferential comprehension difficulties in patients with left frontal lobe impairments. Other researchers like Happe (1994) and Jollife and Baron-Cohen (1999) propose that 'weak central coherence' (the cognitive style of attention to detail rather than extracting the context-dependent meaning from situations) can explain the inferencing and non-literal language comprehension deficits (eg indirect requests, idioms, puns) in people with HFA.

Norbury and Bishop (2002) emphasize that the type of inference required in studies could affect the results. In Bishop and Adams' (1992) study 54% of the inferencing questions required the child to refer to a character's mental or emotional state whereas only 15% of the questions in Norbury and Bishop's (2002) study required such an inference. The children with PLI performed better in the latter study. Consequently, they recommend that future studies include equal numbers of causal and mental state inferences.

1.5.3 The Strange Stories Test

This higher level Theory of Mind (TOM) task was first developed by Francesca Happe (Happe, 1994) as an advanced test of social understanding and theory of mind. It involves the ability to infer the thoughts and feelings of story characters. The original test contained 13 physical state stories and 24 mental state stories. The latter include pretending, joking, lying, telling white lies, figures of speech, misunderstandings, persuasion, contrary emotions, appearance vs reality, forgetting, sarcasm, irony and double bluff (the last two involving third-order theory of mind). The reader is read out a short scenario and asked if the person in the story has said something true and why they said what they did. Examples of two of the stories are given in full below (more examples of the physical and mental state stories from the test are given in Appendix ii:

Pretend (Banana)

Katie and Emma are playing in the house. Emma picks up a banana from the fruit bowl and holds it up to her ear. She says to Katie 'Look! This banana is a telephone!'

1. Is it true what Emma says?
2. Why does Emma say this?

Sarcasm (Picnic)

Sarah and Tom are going on a picnic. It is Tom's idea, he says it is going to be a lovely sunny day for a picnic. But just as they are unpacking the food, it starts to rain and soon they are both soaked to the skin. Sarah is cross. She says 'Oh yes, a lovely day for a picnic alright!'

1. Is it true what Sarah says?
2. Why does she say this?

Detailed scoring procedures are included within the study.

Happe's (1994) study included twenty-six typically developing controls aged 6:06 to 9:07. She found that no groups had difficulty interpreting the physical scenarios, however, the typically-developing younger children and adults with high-functioning autism had difficulty understanding the mentalising tasks in the social inferencing scenarios. The older typically-developing children, on the other hand, were able to understand these.

Jolliffe and Baron-Cohen (1999) replicated Happe's 1994 study with 17 young adults with ASD (mean age 30.7), 17 young adults with Asperger's Syndrome/High Functioning Autism (mean age 27.7), and 17 typically developing controls (mean age 30). The clinical groups made significantly more errors than the control group The story types of joke, figure of speech, irony, and contrary emotions were the most difficult for the high functioning autism group but some of the younger controls also had problems with these story types. They concluded that although the Strange Stories test clearly identifies mentalising deficits in individuals on the autism spectrum, these could be due to an impaired theory of mind or central coherence reasons. They recommended future research in this area to address this issue.

Kaland et al (2005) also replicated Happe's 1994 study with Danish children:

Their study groups were: twenty one children and adults with Asperger's Syndrome of average IQ, aged 10:02-20:04 and twenty age-matched controls, aged 9:06-20:09. They found that the group with Asperger's Syndrome performed less well than the controls on the mental state inferences but matched them on the physical state control task.

O'Hare et al (2009) used selected parts of Happe's 1994 assessment to obtain norms in theory of mind tasks for one-hundred-and-forty 5 to 12 year-old typically-developing children. They found no significant difference in performance between boys and girls. They found a developmental order of understanding in these tasks. Out of a possible total of 24 the following mean scores were obtained: the 5 year-olds 4.67; the 6 year-olds 6.35; the 7 year-olds 10.88; the 8 year-olds 11.65; the 9 year-olds 15.06; the 10 year-olds 15.57; the 11 year-olds 14.68; and the 12 year-olds 19. No child in the youngest age group understood the concept of persuasion and only a few understood sarcasm. The maximum possible score of 24 was not obtained even by the 12 year-olds. O'Hare et al (2009) concluded that assessment of theory of mind can be helpful in the diagnostic process. A limitation of this study is that it did not assess the children's IQ or verbal comprehension. Another frequent criticism of the Strange Stories tests is that they require a high level of verbal comprehension.

1.5.4 Mental state words

Understanding the mental states of others is linked to theory of mind development and is important for the ability to make social inferences. Mental state words allow us to talk about feelings, beliefs, intentions, and one's own and others' internal thoughts. They also help us to understand pretence and deceit. There is a clear link between social cognition, theory of mind and the language of mental state terms. These abilities start to develop around three years of age and continue to develop throughout later childhood and adult life (Spanoudis et al, 2007).

Mental state verbs

These are defined as a class of words that have an internal state as their primary meaning and they require pragmatic inferences about presupposition and implication. Semantically, mental state verbs describe expressions of desire, belief and intention. Pragmatically, mental state verbs help us to understand what is presupposed. Factive mental state verbs (know, realise, to be sorry, to be happy) presuppose the truth of their complement and nonfactive verbs (think) do not. Positive implicative verbs (remember, manage, forget, to be careful) imply that their complements are true, negative implicative verbs (want) imply that their complements are false.

Mental state verb understanding in brain injured children

Eisele et al (1998) tested twenty-four children 4-17 year–old children with brain injuries on their ability to presuppose the truth of factive sentences (eg Max knew that he locked the door) and to infer the truth or falsity of implicative sentences (eg Max remembered to lock the door). Fourteen children had unilateral left hemisphere (ULH) damage, ten had unilateral right hemisphere (URH) damage and fourteen were typically-developing age-matched controls. The children with ULH damage had difficulty understanding presupposition and implicature. The children with URH impairment, on the other hand, had difficulty understanding implicature, but not presupposition. These findings give further support to the evidence for inferential comprehension difficulties in children with brain injuries outlined in section 4.

Mental state verb understanding in children with High Functioning Autism (HFA)

Children with HFA have been shown to have difficulty understanding and using mental state words, in particular mental state verbs. Dennis et al (2001) contrasted the ability of 8 children with HFA (each with a verbal IQ of over 70 and a performance IQ of 100) and typically-developing children to understand and use: pragmatic inferences about given or presupposed knowledge in mental state words, bridging inferences, elaborative inferences involving figurative language, and the intentional inferences involved in speech acts. The children with HFA were able to: make inferences from mental state verbs to given or presupposed knowledge, but were not able to infer what these verbs implied in context; understand elaborated or intended meaning; or make contextual inferences about thoughts. They also failed to make inferences about social scripts. These abilities are crucial for successful social communication. Criticisms of this study are that it was an extremely small scale study and some of the children had a potential verbal-nonverbal IQ gap of 30, which suggests a significant language impairment in those children.

Mental state verb understanding in children with pragmatic language impairment (PLI) and specific language impairment (SLI)

Spanoudis et al (2007) tested 86 Cypriot children in grades 3-6: 18 children with PLI (mean age 11.72), 28 children with SLI (mean age 8.82) and 40 typically-developing children (mean age 12.18) on their understanding of mental state verbs. They used a range of mental state verbs (compare, guess, decide, believe, agree, think) in non-inferential (semantic, give a synonym for) and inferential (pragmatic, set in a story context) tasks. Both factive (know, remember, understand, learn, forget) and non-factive (promise, agree, think, imagine, believe) mental state verbs were used. Their results demonstrated that the children with both types of communication impairments found both types of mental verb tasks significantly more difficult to understand than the typically-developing children and the children with PLI were more impaired understanding the inferential mental verbs than the SLI children. The authors reported that The Children's Communication Checklist (Bishop, 2003) used in conjunction with mental verb measures classified the three groups well. A criticism of this study is that the mean ages of the groups are not comparable, with the typically-developing children having the highest mean age. This could have skewed the results. Also, the findings from Greek mental state verbs might not apply to English mental state verbs.

1.5.5 Faux pas

Verbal faux pas occur when a speaker unknowingly says something that upsets someone else. Pearson and Pillow (2016) state that the ability to understand verbal faux pas is a measure of advanced social understanding that develops between the ages of 7-11 years of age. Comprehension of faux pas requires the ability to take someone else's perspective, understand the link between beliefs, emotions and intentions and to have beliefs about someone else's belief (recursive mental states). Baron-Cohen et al (1999) assessed fifty-nine 7, 9 and 11 year-old typically-developing boys and girls and 12 children with High Functioning Autism (HFA) or Asperger's Syndrome (AS) on a faux pas story task. The children were asked a detection, identification, comprehension and false belief question for each story. An example of one of the 10 scenarios with two types of questions is given below:

Sally has short blonde hair. She was at her Aunt Carol's house. The doorbell rang. It was Mary, a neighbour. Mary said "Hello", then looked at Sally and said "Oh, I don't think I've met this little boy. What's your name?" Aunt Carol said "Who'd like a cup of tea?".

Comprehension question: Whose house was Sally at?

False belief question: Did Mary know that Sally was a little girl?
The results showed that performance on the task improved from 7 to 11 years of age in the typically-developing children and that the girls did better than the boys.
The mean scores out of a possible 10 for the typically developing children were:
7 year-olds: Boys – 2.9 / Girls – 3.8
9 year-olds: Boys – 4.6 / Girls – 7.2
11 year-olds: Boys – 7.9 / Girls – 8.5
The children with HFA or AS or HFA were impaired on the task compared to the controls. However, some children in the clinical group were able to understand the faux pas but they still made faux pas errors in their own social communication. They recommended that future research should focus on faux pas production rather than comprehension.
In a follow up study, Banerjee and Watling (2005, in Pearson and Pillow, 2016) asked 6 and 9 year-old typically-developing children to identify the listener's feelings about what the speaker had said (happy or sad) and whether the speaker meant to offend the listener. The 9 year-olds performed better than the 6 year-olds. They were all better at answering the comprehension questions, identifying the faux pas comment and judging the listener's emotional state than they were at working out the speaker's intention and ignorance. Faux pas comprehension is better when children are accepted by their peers and worse when they have social anxiety.
Pearson and Pillow (2016) investigated children's and adults' understanding of faux pas, insults, and apologies. Like faux pas, insults and apologies are also social acts that involve the thoughts and feelings of speakers and listeners. Insults, which are deliberately and explicitly negative, are usually understood by 7 years of age. Sixty-four children, aged 7 to 11 and sixteen adults (mean age of 19) were presented with brief stories in which the speaker either committed a faux pas or made an insulting remark, but then apologized to the listener. An example of an insult story, based on the Sally faux pas story is given below:
Sally is a young girl with short blonde hair. She was at her Aunt Carol's house. The doorbell rang and her aunt Carol answered it. It was Nate, the boy next door, who Sally had not met. "Hi", Aunt Carol said, "Nice of you to stop by. This is Sally my niece." Nate said "Hello", then looked at Sally and said, "Your niece looks like a boy."

Did Nate know Sally was a little girl?

Who is Nate?

Had Nate met Sally before?

The results indicated that the 9, 11 year-olds and adults all age groups were able to work out that the speaker did not know (faux pas) or did know (insults) information about the listener and they knew that the speaker would feel worse in the faux pas situation than the insult scenario. The younger children were not able to do this. The children reported that an apology would make the listener feel better more than the adults did. A criticism of the faux pas tests is that they require a high level of verbal comprehension.

1.5.6 Understanding sarcasm and theory of mind in children with head injury

Channon et al (2005) state that closed head injury (CHI) is associated with communication difficulties in everyday social interactions, in particular impaired comprehension and use of sarcasm (where contradictory messages are given simultaneously with verbal and body language). They also report findings that children with CHI have significant difficulties with a range of mentalising tasks, indicating that they have impaired theory of mind. Mentalising stories were used to assess theory of mind and physical stories were used as a control. Examples of both types of story are given below:

A physical story

Kenneth grew vegetables in his garden. There were rows of carrots, potatoes, and cabbages. One morning he went to pick a cabbage. Wire fencing protected the vegetable patch. The metal had rusted and there was a small hole in the fencing. Kenneth walked over to the cabbages. There were no cabbages left in the patch.

Question: Why were there no cabbages left?

A mentalising story

Dave wanted to impress his new girlfriend Marie. He was cooking her a meal, but had never cooked before. Marie hoped it would be successful. Dave told her he had spent all day preparing it. When it came out of the oven it was badly burnt. Marie ate all her meal Afterwards she took a second helping of the food.

Question: Why did Marie take a second helping?

The CHI group was selectively impaired on the mentalising component of this task, and mentalising scores correlated with sarcasm comprehension. These findings link into the inferential comprehension difficulty results of children with brain injuries outlined in section 4. A criticism of tests of sarcasm is that it is difficult to ensure that the body language and intonation features contradicting the verbal message will be consistent across different testers.
Burdon et al (2016) evaluated the use of The Awareness of Social Inference Test (TASIT, McDonald et al, 2003), which is an assessment standardized on an Australian population, with young and old British adults ranging from 18 to 90 years. The TASIT assesses understanding of spontaneous emotions, sincerity, sarcasm and lies via the participants watching videotaped vignettes of everyday social interactions. It was originally designed to be used with young Australian adults who had sustained a severe Traumatic Brain Injury (TBI). Social perception difficulties are commonly seen in people with TBI, dementia and right hemisphere stroke (Burdon et al, 2016). Burdon et al found that the normative data for the younger British adults were similar to that of the younger Australian adults, but the normative data for the older Australian adults were not similar to that of the older British adults.

The previous sections have defined the terminology used, justified the need for an assessment of inferential comprehension and idioms, outlined the normative data on typical development of inferential and idiom comprehension and the difficulties many children with communication impairments have in these areas of verbal comprehension. The diagnostic and methodological complexities of assessing these areas of language were also summarized.
The next section (6) goes on to describe the process of creating a formal language assessment in more detail.

CHAPTER ONE SECTION SIX: 1.6 Creating a language test

1.6.1 Formal test development

1.6.1.1 Test definition

McCauley (2001, p49) defines a test as:

'a behavioral measure in which a structured sample of behavior is obtained under conditions in which the tested individual is expected (or at least has been instructed) to do his or her best.'

Murphy and Davidshofer (2005) define a psychological test as a sample of behaviour which is obtained under standardised conditions and for which there are established rules for scoring and obtaining quantitative information.

Psychological, educational and language tests started to be developed at the turn of the 20th century. Speech and Language Therapy is a relatively new profession, with the UK professional body only being established in 1945 (Robertson et al, 1995, Stansfield and Armstrong, 2016) so specific communication assessments have been evolving since then.

1.6.1.2 A brief history of psychological tests

Formal psychological tests were devised just over one hundred years ago (Gregory, 2004). The first modern intelligence test, developed in Paris by Alfred Binet and his colleague Theodore Simon in 1905, was designed solely to identify those children who needed special educational provision. The Binet-Simon Scale involved both verbal and nonverbal/performance tasks. The verbal tasks included responding to 25 abstract comprehension questions such as: 'When a person has offended you and comes to offer you his apologies, what should you do?' The 1908 revised version of the scale included explanation of verbal absurdities, some of which were macabre (eg 'The body of an unfortunate girl was found, cut into 18 pieces. It is thought that she killed herself'). French children were amused by this question but American children found it distressing, illustrating the importance of taking into account cultural factors when designing a test.

The 1908 scale was standardised on 300 typically developing children aged 3-13 years. All items passed by 80-90 percent of a year-group were placed within that age band. Binet and Simon also introduced the concept of a baseline to begin testing. A child’s basal age was set as the age level at which not more than one test item was answered incorrectly. However, Binet always emphasized that children’s mental level as calculated from his test should NOT be taken as an absolute measure of their intelligence. He wanted the test results to be used as a basis for providing help for the child (McCauley, 2001). He also pointed out that the severity of learning delay is relative to the child’s chronological age. For example a 5 year-old with a 3 year learning delay is much more impaired than a 13 year-old with a 3 year learning delay. Consequently in 1916, Terman and his associates at Stanford translated the test from French into English and produced the Stanford-Binet version of the test. This enabled an intelligence quotient to be computed from the mental age divided by the chronological age and multiplied by 100 to remove fractions, giving a better measure of the relative performance of children compared to their peers. The term IQ emerged at this point. However, Simon later described the concept of IQ as a betrayal of his and Binet’s original humanistic scale-development objectives (Gregory, 2004).

Results of early IQ tests were sometimes used abusively. For example: Goddard’s recommendation that the 3% of children with an IQ delay of 4 or more years (described at that time as ‘feebleminded’) be segregated from society for fear of contaminating the rest of society, mass IQ testing of immigrants to America arriving on Ellis island in the early 1900s to justify refusing entry to low scoring individuals, limiting entry to the armed services, and using IQ scores to segregate certain individuals in prisons (McCauley, 2001).

The Stanford-Binet IQ test was used for decades and revised as recently as 2003. However the Wechsler Intelligence Scales (first developed in 1949) provided both separate Verbal and Performance IQ scores as well as a full scale IQ score, whereas the earlier versions of the Stanford-Binet only provided an overall IQ score. Consequently the Wechsler test began to take over in popularity from the Stanford-Binet in the 1960s.

1.6.1.3 Aptitude tests

Aptitude tests differ from intelligence tests in that they measure just one ability domain, or several abilities in that one domain. A new statistical technique, factor analysis was developed (Spearman 1927, in Gregory, 2004) to identify which aptitudes were distinct from each other. Thurstone (1938, in Gregory, 2004) defined the specific factors of primary mental ability as being: verbal comprehension, word fluency, number facility, spatial ability, associative memory, perceptual speed and general reasoning. Aptitude tests have been developed to probe specific areas in more depth than in general IQ tests. For example, pilots have to pass a general IQ test but then more detailed aptitude tests before they are taken on for training. Kline (2000) states that 'aptitude' can be a confusing term. It usually refers to a collection of abilities that are of value in a particular culture. He confirms that verbal assessments are aptitude tests. The new test designed in this research is an aptitude test covering the domains of verbal comprehension and reasoning (specifically understanding inferences and idioms).

1.6.2 Current psychological tests

1.6.2.1 Definition of a test

These definitions apply to psychometric tests;

- McCall (1939) stated that anything that exists in a certain amount can be measured. A psychological test must lead to a score or categories which can measure a person's performance in some way.
- Gregory (2004, p30) defines a test as: 'a standardised procedure for sampling behaviour and describing it with categories or scores'.

Most tests are carried out in a standardised way, measure different kinds of behaviours which are scored or categorised, have norms or standards, and can be used to predict other non-test behaviours. The 8 main types of psychological tests currently in use are: Intelligence Tests, Aptitude Tests, Achievement Tests, Creativity Tests, Personality Tests, Interest Inventories, Behavioural Procedures and Neuropsychological Tests (Gregory, 2004).

1.6.2.2 Standardisation of tests

Standardisation is an essential feature of psychological testing. A test is said to be standardised if the procedures for administering and scoring it are consistent across testers and settings. It depends on having a competent examiner and clear and comprehensive directions for administering the test (usually provided in an accompanying manual) so that the test procedures can be replicated exactly by a range of testers. A good test manual will include standardisation, norms, administration, reliability and validity information.

As tests are analysing an external sample of behaviour to make inferences about unobservable characteristics, the results will always have a degree of imprecision and measurement error.

Practical time constraints mean that a test represents only a representative sample of behaviour under scrutiny as a whole population cannot be tested. However, a specific sample allows the examiner to make inferences about this domain and other domains (eg poor performance on a specific picture naming test indicates that the person being tested has a poor general expressive vocabulary).

1.6.2.3 Criterion-referenced and norm-referenced tests

Criterion-referenced tests do not compare individuals with a reference group, but measure where they stand compared to, for example, specific educational or speech and language criteria (eg has accuracy in a specific arithmetic task, performance in phonology assessments).

Norm-referenced tests make up the majority of formal tests and do compare an individual's performance to the standardised sample. Test developers usually produce norms which are a summary of the test results from a large representative group of the population being tested. Norms establish mean performances for different groups and show how frequently scores deviate from these means. The ultimate goal of an assessment is to predict additional behaviours other than those directly sampled in testing, which requires post-test validational research once the test has been released. However, it cannot be guaranteed that a published test measures the characteristics it claims to measure.

1.6.2.4 Test versus assessment

The terms test and assessment are often used interchangeably, but they have slightly different meanings. A test is one small specific source of information that forms part of the wider assessment process which aims to collate all relevant information from a range of sources about a person (qualitative as well as quantitative). The compiled information then allows inferences about characteristics to be drawn and predictions about specific behaviours to be posed. Gregory (2004) emphasizes that assessment is 'an inherently subjective process that requires the examiner to sort out conflicting information and make predictions based on a complex gestalt of data.' (p33).

Combined test results and assessment information are used to help make a differential diagnosis of, for example, educational difficulties and/or medical conditions. A proper diagnosis supplies information about: the possible cause of a person's condition, their strengths and weaknesses, and the best choices for appropriate support and intervention.

Tests can be administered to individuals or whole groups of people at the same time.

1.6.2.5 Test uses

The most frequent use for a psychological test is to help professionals make decisions to help a person, in partnership with them and key people in their lives. For example: type of educational provision, admission to higher education, employment offers. However, there are five main uses of tests: classification; diagnosis and treatment planning; self-knowledge; programme evaluation; and research.

The assessment developed by this research is a norm-referenced individual test that will form part of a child's assessment for communication impairments. It aims to assist in differential diagnosis and treatment planning and in future research in this area.

Open tests are available for anyone to administer without formal training (eg the Bracken School Readiness Test, Bracken, 2002). Closed tests can only be carried out by qualified people who have been trained to use them (eg The New Reynell Developmental Language Scales, Edwards et al, 2011).

The reasons for the strict restrictions on closed tests are that unqualified testers could actually cause some harm, and not following exact test procedures or leaking test content to the general public can invalidate the results.

1.6.2.6 Factors affecting the testing process

A test needs to be as well-designed as possible but external factors are equally important to ensure accurate test results. The manner of administration; tester characteristics; the testing context; the experience and motivation of the tester; and the scoring method are all of crucial importance (Gregory, 2004).

Early developmental and language research into children was carried out in a very artificial clinical setting which did not elicit natural behaviour or communication from the child. In response to this Bronfenbrenner and Morris (1998, in McCauley, 2001) created a bioecological model of development to take into account contextual factors and characteristics of the child. The child's environment, particularly the social environment, is of crucial importance to the child's development. Family, language, culture, society and the clinician's context shape the design and validity of tests for children. Effective assessment and intervention for children needs to include collaboration with their family members. The interaction of the tester with the family needs to be positive, inclusive, nonintrusive and respectful of the family's culture and values. Parents are vital to give the details of and background to the child's communication difficulties. They provide answers to screening assessments and they may even help to administer some tests.

Cultural factors can hugely affect the clinician's interaction with the family and the child. McCauley (2001) outlines just a few of these, but it is a rapidly expanding area of research. There are major differences across cultures in child-rearing practices: who the main decision maker in the family is, expectations around additional language or dialect use in different contexts, and differences in how communication impairments are perceived and explained. It is crucial to be aware of and take into account any relevant cultural considerations when assessing and providing intervention for children. McCauley (2001) emphasizes that, regardless of culture, the clinician should find out what a family's attitudes are and how it functions.

Murphy and Davidshofer (2005) describe early attempts to make 'culture-free' tests. However, it has been found to be an impossible task, as there will always be some cultural influences in assessments. As learned behaviour and many communication skills are a function of culture, no test can be completely culture-free. Better terms would be 'culture-fair' or 'culture-reduced' for less culturally biased tests and 'culture-loaded' for more culturally biased assessment. A culture-reduced test uses universal objects, symbols and information. However, researchers often disagree about what constitutes a culturally-fair or culturally-loaded test. Murphy and Davidshofer (2005) list some characteristics of culture-loaded tests as: paper-and-pencil tests, reading required, written response required, speed tests, verbal content, and recall of past-learned information. Some characteristics of culture-reduced assessments are: performance test (the subjects answer as many questions correctly as they can), purely pictorial, oral response, power tests (not timed), nonverbal content, solving novel problems.

The only obviously culture-loaded aspect of the current researcher's newly devised test from the above list is that it is verbal (which it has to be as it is a test of language comprehension). The culture-reduced characteristics are: it is a performance test, the required answers are oral, the responses are not timed, and there is an element of solving novel problems. So the new test contains 1/6 culture-loaded elements and 4/6 culture-reduced elements. This would indicate that the test is nearer the culture-reduced end of the cultural bias continuum. The other considerations of cultural factors in the test are the choice of scenarios that are deliberately not culture-specific and the choice of children and adult's names in the assessment. Overall it has been designed to be, as far as possible, a 'culture-fair' assessment.

Contextual factors affecting SLTs' practice with children include following national and international advisory and legal guidelines. For example: using the diagnostic criteria for specific conditions from the American Psychiatric Association Diagnostic and Statistical Manual version 5 (DSM-5, APA, 2013) or the European World Health organisation (WHO) International Classification of Impairments, Disabilities and Handicaps (WHO, ICD-10, 2004), adhering to the rulings of the disability Equality Act (Gov.uk), and implementing the National Institute for Health and Care Excellence (NICE) guidelines for certain client groups. Other factors affecting SLT provision include: complying with local policies and procedures, staffing needs and shortages, caseload numbers and complexity, the expanded role of the SLT over the last three decades, and increasing numbers diagnosed of certain client groups. For example, the diagnosis of people with ASD has increased significantly in the last decade. It is now seen in 1.1 per cent of the population and affects seven times as many males as females (Brugha et al, 2012).

1.6.2.7 Standardising the test administration

Testers should follow the exact procedures prescribed in the test manual to ensure the reliability of the test results. The American Psychological Association's revised Standards for Educational and Psychological Testing (AERA 2014, in Plake and Wise, 2014) emphasize that the specific instructions to test takers, time limits on test items, how test items are presented, accepted forms of response, and supporting materials or equipment all need to be adhered to exactly. These standards represent the best professional judgements on how to design, develop and use tests that provide high levels of reliability, validity and fairness (Plake and Wise, 2014). McCauley (2001) highlights that the stance of the AERA over the previous 2 decades had shifted from a 75% emphasis on test standards and 25% emphasis on text administration to a 40% emphasis on test standards and a 60% emphasis on test use, indicating the importance of the test user in measuring the quality of the assessment. She also states that many ethical issues need to be considered in test administration.

In language tests it is very important not to give extra verbal or nonverbal clues (eg telling the child another similar word in a vocabulary test, or using intonation and facial expression to support the meaning of idioms) as this can invalidate the test. However, it is also important to include some flexibility in procedures, based on what exactly is being assessed and on the tester's professional judgement. Gregory (2004) gives an example of a nervous college student initially interpreting a question very literally. The examiner asked 'How much is four dollars and five dollars?' The student's reply was: 'Four dollars is four dollars and five dollars is five dollars'. When the examiner repeated the question, this time emphasizing the AND, the student realised it was an arithmetic question and answered it correctly. If the assessment was a test of nonliteral comprehension, however, this repetition with new extra emphasis would not be allowed. Gregory (2004) assures us that minor adjustments in test procedures regularly occur and such flexibility is essential to obtain valid results. Examples of the need for flexibility in accepting responses from the assessment being developed in the current project are: joke answers, and dialectal variations – examiners are advised to use their clinical judgement in deciding if an answer is acceptable or not. If the essential standardised procedures are not adhered to (eg allowing patients with expressive aphasia more than the specified time to respond in a test), the examiner must state this in any written report and the test results can only be used to provide qualitative not quantitative information (ie the calculation of a standard score would be invalid).

Essential features of successful test administration include: the examiner being very familiar with the test materials and administration procedures before carrying out the assessment for the first time, making allowances for certain disabilities (visual, hearing, communication and motor impairments etc) which are permissible within the test procedure framework, the examiner establishing a good rapport with the testees, particularly essential if they are children, and reduction of anxiety of the testee as much as possible. McCauley (2001) states that the clinician plays a critical role in assuring test validity through accurate administration procedures and establishing a positive rapport with the examinee.

Test scoring errors are very rarely the result of individual tester differences, as scoring criteria for tests are usually very detailed. If any scoring errors occur they are much more likely to be clerical (eg noting down or adding up the marks incorrectly or using the wrong table to interpret the results).

1.6.2.8 Norms

Tests need to be standardised against an appropriate norm group so that the individual test scores can be interpreted correctly. The test results need to be consistent and replicable (reliability) and measure what they are supposed to measure (validity).

In most tests raw scores need to be converted to a derived score based on the norms from the population assessed. The norm group is the sample of participants who represent the population for whom the test has been designed. The purpose of standardization is to obtain the distribution of raw scores in the tested population so that derived scores (norms) can be drawn from these and published. These norms indicate how well an individual does compared to his or her peers. Norms include age equivalents, standard scores, and percentile ranks. Norms quickly become outdated so they need to be regularly updated.

1.6.2.9 Scoring types

Age equivalents are the mean (average) score for that person's age. Average scores mean that some participants score below and some score above average. Therefore not every person who scores below their age equivalent score has a problem. McCauley (2001) emphasises that age equivalent scores are less reliable than other types of scores. This is best illustrated by tests on children, where the significance of the result depends on how the mean score changes with age and how much variation in test scores is typically found at a given age (Bishop, 1997). Consequently, percentile and standard scores are more reliable than age equivalent scores.

A percentile score illustrates the percentage of people in the standardised sample who score above or below a particular raw score. The percentiles range from 0 to 100, with 50 being a mean percentile.

A standard score represents the participants score in terms of how far away from the mean it is in standard deviation units. The formula for this is: $z = (X-M)/SD$.

1.6.2.10 Test construction

Gregory (2004) states that the development of a new test evolves slowly over time and is both a science and an art. McCauley (2001) describes the test construction process as six overlapping stages: planning the test, writing possible test items, conducting an item try-out, conducting an item-analysis, obtaining norms, and collecting validity and reliability. Gregory (2004) adds in a final stage of publishing the assessment. When defining the test the developer must be clear about what exactly it will measure and how it will differ from existing tests in the same field. There are always new research findings that contribute to the amendment of existing assessments or the development of new ones. The four main scaling measurements are: nominal (naming or categorising), ordinal (ranking), interval scales, and ratio scales. Most psychological tests use interval scales even though it is difficult to show absolute equality of intervals for such assessments (Gregory, 2004). Most standardised tests of children's language use interval scales (McCauley, 2001). A pilot study on a small population is usually conducted in the early stages of testing. More items than needed for the final assessment are trialled and then the number is revised down using item analysis. This can include the use of item-difficulty index, item-reliability index, item-validity index, item-characteristic curve, and an index of item discrimination. The type of item selection tests selected depend on the type and purpose of the assessment. Once unsuitable items have been excluded from the test the revised test should contain fewer, but more discriminating items than the original version and should be more reliable and accurate. This new version then needs to be trialled on a similar but much larger population. Ideally the test revision process should include feedback from the examinees on how they think the test could be improved, although this is not always appropriate for some testing (eg on young children). The final stage of test development includes finding a publisher and producing user-friendly test materials, a technical manual and a user's manual. The final test must be as quick and as smooth as possible to administer. It may help to have some of the test instructions written on the test form or on the back of an easel-style test booklet. The AERA test standards (Plake and Wise, 2014) recommend that technical test manuals describe the rationale and recommended uses for the test; provide specific cautions against anticipated misuses of the test; cite representative studies regarding general and specific test uses; identify any special qualifications needed to administer and interpret the test; provide revisions, amendments and supplements as needed; use promotional material that is accurate and research-based; cite quantitative relationships between test scores and criteria; state if any alternative modes of response are acceptable; provide appropriate interpretations for the test taker; give evidence for the validity of any automated test scoring; and include the essential data on reliability and validity. Gregory (2004) concludes that the development of a new test is a hugely time-consuming and expensive process. It often becomes a multi-million dollar project as it requires a large staff of test construction experts working for several years. Test publishers are consequently very conservative about introducing any new tests unless they offer a very significant improvement on existing assessments.

1.6.3 Developing language assessments for children

McCauley (2001) stresses that language assessments are just part of the clinical decision-making process for Speech and Language Therapists (SLTs) when working with children with language impairments. Language assessment results can help towards solving the puzzle of the range of difficulties with which these children present. They often form part of a wider multi-disciplinary assessment working towards a differential diagnosis. Measurement of language is extremely complex. Assessment of language has direct roots in psychological and educational testing but may also draw on linguistic and acoustic measurements. Measurement of any specific language difficulties is the first step towards effective clinical action to help to improve them. McCauley (2001) describes five different kinds of clinical decision making in managing children with language impairments: screening for a language disorder, diagnosing a language disorder, planning for the management of the language disorder, assessment of change in language skills over time, and identification of the need for additional information in a related area of communication. Standardised norm-referenced or criterion-referenced tests (alongside other procedures) are usually required in all but the first stage. The specific standardised assessments selected depend on the age and profile of the child. These assessments need to be administered in a competent and standardised way to ensure that the results are as accurate as possible. These test results then form part of all information collated by the SLT from a range of other sources and contexts. This

collated information allows the SLT to form hypotheses about the child's specific profile of strengths and needs and to develop a clinical intervention plan.
Testing should be a collaborative process: factors affecting validity when testing children include motivational and enabling behaviours. Children expect the examiner to interact with them in a natural way and they will usually try to perform their best to please the examiner. To gain the children's full co-operation they should be given clear information about the reason for and the content, order and timing of the assessment. Children require the pre-requisite skills to perform on certain tests. For example for verbal language tests they require good hearing, vision and adequate motor skills to respond appropriately. When making decisions on behalf of children, including how best to assess them, it is crucial to take into account characteristics of the child and of the context in which the decision-making is taking place (McCauley, 2001). Andersson (2005) points out that published reviews of language assessments for children are of limited use in helping the clinician's decision making process, as they are usually published well after the test is available for clinical use. She suggests, therefore, that clinicians should possess the tools and skills needed to evaluate the adequacy of language tests for specific purposes (e.g. screening versus diagnostic assessment) and with specific populations of children.

1.6.4 Socio-economic status and language development

There is a large body of research linking lower socio-economic status with poorer language skills (Hoff, 2006). Young children from disadvantaged backgrounds are at higher risk of having delayed speech and language development than their peers from more advantaged socio-economic backgrounds (Basit et al, 2015). Most of the research examines this effect in children in the early years. Some researchers report that these differences in language development increase with age and continue beyond the early years (Vasilyeva et al, 2008). Shore (2015) found differences in expressive language in older primary children, with the children from disadvantaged socio-economic backgrounds performing worse than the children from more advantaged socio-economic backgrounds. There is also a lot of research evidence linking lower socio-economic status to reading comprehension problems and to lower educational attainment. Snow et al (2007, in Spencer et al 2016) followed up children from low-income households in the United States from three years of age until adolescence. They found significant relationships between language, literacy and educational achievement throughout this time period. However, by adolescence the link between language and educational outcome had reduced. Spencer et al (2016) carried out a study examining the link between spoken language, socio-economic status and educational outcomes in 151 British 13 and 14 year-olds who were followed up at 16 years of age. They concluded that language ability, particularly vocabulary, played a more important role in predicting educational outcome than socio-economic status. There appear to be no research studies looking specifically at the link between inferential comprehension and socio-economic status. The examination of this relationship in this current study is therefore justified.

This section has given a brief history of psychological and language testing and has summarised the main requirements for designing a new language assessment. The next section critically examines the current formal and informal language tests that assess aspects of inferential and idiom comprehension.

CHAPTER ONE SECTION SEVEN: 1.7 Current formal and informal assessments that include inferential and idiom comprehension

A review of existing assessments that include inferential and idiom comprehension is vital before a new test focusing on the gaps in these areas can be devised. The section below summarises and critically evaluates these existing assessments. Reliability and validity data is included if it is given in the test manual but many assessments do not include this information. A chart summary of the advantages and disadvantages of each assessment is given at the end of the section on inferential comprehension assessments and at the end of the section on idiom comprehension.

1.7.1 Assessments of inferential comprehension

1.7.1.1 The Preschool Language Assessment Instrument (PLAI) (Blank et al 1978) and the 'Blank Language of Learning Model' (Blank et al, 1978a, Blank et al, 1978b)

This assessment framework was developed in the USA but is now widely used in Britain and Australia. Although old references, they are seminal works and are still very relevant for assessing higher level comprehension in particular. Blank et al (1978a) recorded and analysed the language typically developing pre-school children understood and used for learning in the classroom (the 'language of instruction'). Their 'Pre-school' group included children up to 6 years of age, so it is relevant for children in the Foundation Stage and Key Stage 1 in England. The British Elklan training programme and Language Builders series of books (Elks and McLachlan, 2005) have increased the use of this assessment framework in the UK in the last decade.

The PLAI assessment included: the book, *The Language of Learning - The Pre-school Years,* which details why and how the PLAI was developed; a manual containing pictures (black and white line drawings) and verbal stimuli; and forms for recording the child's answers.

In the Blank model there are 4 levels progressing from concrete comprehension to abstract comprehension based on the degree of distance between the material available to the child and the language the child has to understand and use in dealing with the material. There are 15 test items at each of the 4 levels, so 60 questions in total.

The first two levels, 'matching perception' and 'selective analysis of perception', ask questions where children have to use language to describe or respond to 'here and now' or just-completed experiences. For example:

Level 1

Find one like this; show me what you heard; what did you see?

Level 2

Find the one that can *Verb* (function of object); what's happening? (describing a scene); who, what, where questions; find the one that is X and Y (adjectives); how are these different?; tell me something that is a Z (category).

Levels 3 and 4 are outlined in more detail as this is where verbal reasoning skills begin.

Level 3

In level 3, 'reordering perception', children need to evaluate material and ideas with some restrictions. Language can no longer be mapped straight onto perceptions, it has to be restructured. Some questions require the exclusion of features. The child also has to use 'metalinguistic skills' (ie be able to use language to talk about language). The authors (Blank et al 1978a) give fourteen detailed examples of level three questions:

A. Scan for an object and integrate visual with verbal information ('Find one to use with this') eg Hold up an apple and ask 'Find something we could cut this with.'
B. Describe events following a scene ('What will happen next?')
C. Assume the role of another person ('What would he say?') eg Show a picture of a boy with one shoe on and say 'A boy came to school with only one shoe on. What did the other children say? What did the boy say?'
D. Follow a set of directions ('Do this, then this') eg 'Put on your shoes and get your coat, then meet me at the car'.
E. Arrange pictures in the correct sequence ('Make these into a story').
F. Formulate a set of directions ('Tell me how to _____') eg 'How do you make a jam sandwich?'

G. Formulate a generalization about a set of events ('What happened to all of these ________?') eg add different food colouring to glasses of water. 'What happened to each of these that was the same? (Answer – the water turned into a different colour).
H. Formulate a statement to unify a sequence of pictures ('Tell this story') eg pictures of a boy riding his bike, falling off his bike and holding his bleeding knee and crying. Ask the child to put the pictures in the correct order then tell the story.
I. Concepts: Identify similarities ('How are these the same/alike?') eg show a picture of a horse and a cow. The answer will need to contain information about category and function as well as description.
J. Concepts: Select an object by exclusion. eg Show a few blue items and 1 or 2 other coloured items. Ask 'Find one that is not blue'.
K. Concepts: Select a set of objects by exclusion ('Find the things that are not ....') eg In the bedroom 'Find some things we don't wear'.
L. Concepts: Give an example by excluding a specific object ('Name something that can X but is not a Y') eg 'What has 4 legs but is not a dog?'
M. Concepts: Give an example by excluding a class of objects ('Name something that is not a Y') eg 'I went to the zoo and saw something that was not an animal
N. Concepts: Define words ('What is a _____?')
O. Unusual imitations ('Say this _______') eg Adult whispers a message to the child but asks him to repeat it in a loud voice. This requires metalinguistic ability.

<u>Level 4</u>

In level 4, 'reasoning about perception', includes complex problems that require reasoning about what will or could happen. Reasoning about perception involves: identifying the cause of events, predicting the effects of certain actions, justifying responses, and evaluating an object's essential and nonessential attributes.

Blank et al (1978a) give fourteen detailed examples of level four questions:

A. Predict: Change in position ('Where will .....?') eg Put a doll on a chair. Ask 'Where would she be if she fell from the chair?

B. Predict: Change in structure ('What will happen if .....?') eg Make a pile of 4 blocks. Ask 'What will happen to the pile of blocks if I take this bottom one away?
C. Justify a prediction ('Why will ......?') Show an empty bowl and a clear bag full of marbles. Ask 'If this bowl was filled all the way up with play dough, could I pour the marbles inside it? Why not?
D. Justify a decision: Essential characteristics ('Why wouldn't it ....?') eg Show a rectangular sponge. Ask 'If the sponge were made of this (pointing to a paper triangle) and not this (pointing to a sponge triangle), would it still be a sponge. Why not?
E. Justify a decision: Nonessential characteristics ('Why would it .....?') eg Making popcorn. Ask 'If I had shaken the kernels in a bowl, would they have popped? Why not?
F. Identify the cause of an event ('What made it happen?') eg Knock a plastic cup off the table and ask 'How did the cup get on the floor?'
G. Formulate a solution ('What could you do?') Point to a picture of a hot bowl of soup Say 'This soup is too hot to eat now. What could we do to cool it down?'
H. Formulate a solution from another's perspective ('What could she do?') eg Show pictures of people experiencing difficulty (eg a a girl who has spilled her milk all over the table). Ask them to tell you what is wrong/what the problem is. Then ask 'What could he/she do?
I. Select the means to a goal ('What could we use?') eg Pour beads through a funnel then ask 'Which one of these things could I use to stop the beads from going through the funnel (choice of a button, toothpick, cork, paper clip, safety pin).
J. Explain the means to a goal ('Why should we use that?'). As above – but ask 'Why should we use the cork?'
K. Explain the construction of objects ('Why is ___ made of that?') Why are wellies made of rubber or plastic and not paper?
L. Explain an inference drawn from an observation ('How can we tell?') eg Show a picture of a tin with a picture of a cake on it. Ask 'How can you tell that there is cake in the tin?'

M. Explain the logic of compound words ('Why is this called _____?') eg Why is this called – a keyhole/a buttonhole/paper towel/rain coat.
N. Explain the obstacles to an action ('Why can't we ______?') eg A closed clear box with buttons in it. Ask 'Why can't we touch the buttons?'

The PLAI assessment was standardised on a group of 120 American children aged between 36 and 71 months from middle and lower socioeconomic backgrounds. Inter-rater, and test-retest reliability all fell within an acceptable range and the authors attempted to demonstrate content and predictive validity for the test (James, 1986). The test was designed to be administered by a range of professionals: teachers, speech and language therapists and psychologists. Children were tested individually and the test took around 20 minutes to administer. Responses were scored on a 0-3 scale, with 3 being a fully acceptable response. Some questions were multiple choice and marked numerically. Others were open questions that were scored by qualitative analysis. Scoring guidelines were included in the manual. Seven categories of response were included. Adequate responses were labelled 'fully adequate', 'acceptable' and 'ambiguous'. Inadequate responses were labelled 'invalid', 'irrelevant', 'don't know' and 'no response'

Blank et al (1978a) produced means, standard deviations and percentile ranks for each of the 4 assessment levels (although not for the total test) but emphasized that the test was designed to test an individual child's skill in dealing with the language of instruction rather than comparing their performance with that of their peers.

The gross results for Blank et al (1978a) were: 60% of 3 year-olds understood level 1 and 2 questions, 65% of 5 year-olds understood level 3 and 4 questions, and level 4 skills developed between 4:06 and 6 years of age.

James (1986) concluded there were no other formal assessments (at that time) that provided information about children's understanding of language within an instructional setting. However, she stressed that the PLAI should not be used as an instrument to compare children's performance with that of their peers. It was designed to develop a profile of each individual child's ability to deal with instructional discourse at four levels of difficulty and for that purpose it was a valuable assessment tool.

Curtis and Foord (1996) carried out an extension study of the PLAI on 4½ to 5 year-old Australian children. They were concerned that Blank et al (1978) had collected data on only 20 children in any six-month age bracket in the USA, ten of whom they classified as middle class and ten of whom ‘lower class’ (the vocabulary used at the time to describe lower socioeconomic status).

Curtis and Foord (1996) administered the PLAI and the Peabody Picture Vocabulary Test-Revised (Dunn and Dunn, 1981), which is a receptive vocabulary test, to forty-three typically developing Australian children aged 4:06-4:11 from two ‘Priority 1’ and two ‘Priority 3’ preschool centres. The results supported the findings of Blank et al (1978a) on mean scores for the 20 children in this age group. They also showed that the children from the centres in the lower socioeconomic areas obtained lower mean scores than the children from the centres in higher socioeconomic areas in all the PLAI levels and the PPVT; there was no significant gender difference; and there were highly significant correlations (Pearson’s Correlation Coefficient) between the PPVT and the PLAI levels 2, 3 and 4. One of Curtis and Foord’s (1996) recommendations was that teachers should be aware of the need to adjust their level of questioning with young children so that it matched their level of comprehension.

Blank et al (2003) brought out a second edition of the PLAI in 2003 (PLAI-2) for children aged 3 to 5:11. This was standardised on a sample of 463 children residing in 16 states, and included more extensive reliability and validity data. The pictures were updated from black and white to colour with increased gender and ethnic representation. Otherwise the assessment is very similar to the original PLAI.

Over the years there have been mixed reactions to the language of learning model and the assessments associated with it. Elks and McLachlan (2004) advocate that knowledge of Blank et al’s (1978a) Language of Learning Model would benefit everyone who works with children. They describe it as a flexible assessment and management tool, which can help to identify abstract reasoning difficulties. They claim that it is particularly useful for children with verbal comprehension difficulties. They cover this model in all of their training courses.

James (1986) states that the greatest difficulty in using the PLAI is in scoring the children’s responses, particularly at levels 3 and 4 where a wide variety of acceptable responses is possible. Testers have to rely heavily on their own judgment for the qualitatively marked questions, which affects reliability of the scoring system. In addition, the test authors only gave inter-rater reliability for the numerical not the qualitative scoring.

Elks and McLachlan (2004) point out that, whereas in Blank et al (1978a) levels 1 and 2 are in a developmental order of difficulty, levels 3 and 4 are not. Some children have particular difficulties with sequencing and empathy (level 3) but find problem solving and justification (level 4) easier.

The current researcher's criticisms of the Blank (1978a) model are:

- The norms are based on a very small sample (20 per age group).
- Within the same level some language structures are very simple and some very complex (eg following a set of directions (level 3), can range from 'go to the bathroom and get me a towel' to 'go to the utility room and select a can of cat food from the top left hand cupboard').
- 'Why' questions can range from syntactically and semantically simple to complex (eg 'why is the girl crying?' to 'why is the girl with the spotted scarf putting her purse behind the canister?') but these are all included in level 4.
- The difficulty level of describing differences and similarities between objects (levels 2 and 3 respectively) and defining words (level 3) is dependent on the developmental level of the vocabulary used (eg 'banana' is easier to define than 'melt').
- Some level 3 questions require much better expressive language to answer them than some level 4 questions. For example, 'how do you make a jam sandwich?' (level 3) is more complex than 'where would she be if she fell from the chair?' (level 4).
- The PLAI scoring system involves many subjective judgments.

<u>Other assessments based on the Blank et al (1978a) 'Language of Learning Model'</u>

<u>1.7.1.2 The Liverpool Language Screen (unpublished data)</u>

This British verbal reasoning screen for 4-7 year-old children was developed by the Liverpool Language Units SLT team in the 1990s. The standardisation procedure was led by the current researcher.

The rationale for the development of this screen was that many of the children with Specific Language Impairment (SLI) with whom the team worked in the language units scored within normal limits on standardised assessments of verbal comprehension, eg the Test for Reception of Grammar (Bishop, 1989), but still exhibited significant comprehension problems, particularly with verbal reasoning, more abstract language and non-literal language comprehension. This applied particularly to children with receptive or pragmatic language impairments (PLI). There were (and still are) no formal verbal reasoning assessments available for this population of children.

In the 1980s, the department had already developed the ‘Liverpool Language Cards’, an assessment and therapy tool, ranging from single word comprehension and expression to more abstract language comprehension, which was based on the Blank et al (1978) model.

The current researcher used a selection of the higher level questions from the Liverpool Language Cards to make a screening tool which was then trialled on seventy-five boys and seventy-one girls aged 4 to 7 (Key Stage 1) in 6 mainstream primary schools in Liverpool. The questions were verbally presented and required a verbal response. However single words or short phrases were sufficient to express a correct response so only a basic level of expressive language was required. A few questions had supporting pictures, given in brackets below.

An example from each sub-section is given below. The full assessment is given in Appendix iii.

1. Taking Another’s Perspective
2. If you were cold, would you take your gloves off?
2. Predictions
5. What would happen to a flower if it didn’t get any water?
3. Obstacles to Action and Experience
9. Why can’t he put the shoe on? (Picture of a boy trying to put a baby’s shoe on).
4. Cause – Effect/ Prevention
20. Why do people cry?
5. Formulating solutions to a Problem
25. How can you talk to somebody who is hundreds of miles away?

6. Explanation of an Inference

29. How can you tell that the children have been eating biscuits? (Picture of 2 children with an open packet of biscuits and a plate each with crumbs on).

7. Construction of Objects

37. Why are stamps sticky on the back?

8. Story Comprehension (9 questions)

All correctly answered questions scored 1 point except 13, 14, 15, 20, 25, 33, 37, 40, 44, 45, 46, 48 which could score 1 or 2. Marking criteria were agreed upon. The maximum total score was 61. The children were divided into six month age bands. No gender differences were found in the scores so the boys and girls scores were combined.

The total mean test scores were:

Table 1.8 Mean total test scores for the Liverpool Language Screen

Age Group	Total Score
4:00-4:05	39
4:06-4:11	42
5:00-5:05	44
5:06-5:11	50
6:00-6:05	51
6:06-6:11	53

Once these norms from typically developing children had been obtained the screen was used to assess the verbal reasoning skills of children with SLI. Older children (Key Stages 2 and 3) with receptive language difficulties (including children with PLI and ASD) scored similarly or worse than the typically-developing 4 to 7 year-olds in many of the screen sub-tests. They had particular difficulty understanding and responding appropriately to the following questions.

1. If you hurt your friend would he/she be happy?

12. Why can’t he see who he’s caught? (there is a picture of a blindfolded boy grabbing a girl’s pigtail).

25. How can you talk to someone who is hundreds of miles away?

28. How could you find out what your teacher looked like when s/he was a little girl/boy?

Nearly all of the sub-test 6 (explaining inferences) questions (29-36) were difficult for the SLI children to understand, particularly those involving Theory of Mind (eg How do you know X?). These are all Blank et al (1978) level 4 questions.
A positive feature of this assessment is that it used British rather than American vocabulary. However, although it was a tool developed to meet a perceived local need, in retrospect, there were more weaknesses than strengths with this verbal reasoning screen. A strength was that it was standardised on a relatively large population (50 children per year group). The weaknesses were that the sample was from only one geographical area; limited descriptive and no inferential statistics were produced on the data; even the 4 year-olds could easily answer virtually all of the questions in some sections correctly; the story questions turned out to be not very useful as they assessed different areas of comprehension than verbal reasoning and were not always included in the screening assessment, meaning that a total test score was not computable. Despite these drawbacks, the assessment clearly met a need, as it continued to be used as a screen until the TALC1 and TALC2 were developed. The current researcher still uses it to obtain quantitative and qualitative data on children's verbal reasoning skills and it formed the early impetus for this more rigorous study.

1.7.1.3 The Test of Abstract Language Comprehension (TALC) (Elks and McLachlan, 2004)

The TALC is a non-standardised British assessment for 3-6 year-old children and is in 2 parts; a picture assessment and a general assessment. The child is required to demonstrate comprehension by replying expressively to a series of questions linked to pictures.

Part 1: Picture Assessment

There are 6 colour picture scenes with 10 questions per picture which are a mixture of Blank level 1-4 questions. The questions from one of the pictures in the assessment are given below.

Picture 1: Two girls lying on the floor colouring in together

Questions (Blank level is given in brackets):

1. Point to the book (1)
2. Show me a big flower (2)

3. What are the girls doing? (2)
4. This girl (adult points) is using a felt pen. What else can she colour with? (2)
5. How are the girls different? (2)
6. How are the girls the same? (3)
7. How do the girls feel? (3)
8. What are they saying to each other? (3)
9. How do you know they are friends? (4)
10. Why should the girl put the lids on her felt pens? (4)

The general assessment (part 2) gives a list of questions from different Blank levels which can be asked in natural contexts.

If the child achieves 80% or more of the questions at a particular level s/he is said to be competent at that level. 80% scores are: 5 out of 6 for level 1 and 14 out of 18 for levels 2, 3 and 4. There is also a table showing the percentage scores for all levels.

If the child achieves less than 80% in a level, the Elklan authors recommend language intervention at this level until it is consolidated. They suggest using the advice and activities in the Elklan Language Builders and Early Language Builders (Elks and McLachlan, 2004).

The assessment can be repeated and changes in the percentage scores used to monitor progress.

The TALC-2 (Elks and McLachlan, 2010) was developed after the original assessment (TALC-1) to assess and develop the verbal reasoning skills of pupils aged 11 and over. It is also based on the Blank et al (1978a) normative data. It contains parts 1 and 2 similar to the TALC-1 but has a third section which provides examples of how to modify questions whilst teaching different curriculum areas.

A positive feature of the TALC assessments is that they are British and so use British vocabulary. The current researcher's criticisms of the TALC (1 and 2) are: The authors used Blank et al's (1978) normative data to construct their assessment and did not standardise it on a typical British population of 3-6 year-olds. They do not include detailed marking criteria for acceptable or unacceptable answers. In fact they state that, as it is a test of receptive language, as long as the child gives some indication of understanding the question it is marked as correct. Some of the pictures can be interpreted in more than one way, eg picture 2 (party), question 19 'Why is

Jack leaning across the table?' The expected answer is that he is going to blow out his birthday cake candles but he could also be reaching for sausages. Picture 3 (at the park) question 26 'What happened to the baby?' The baby could have fallen over in a puddle or been kicked by the girl on the swing. There are errors with the numbering of the test questions on the form and with the adding up of the total level 3 questions and the total number of test questions. In addition, in the current researcher's clinical experience, practising SLTs only use the picture assessment questions not the general assessment questions due to the time these take to administer.

The TALC-2 is designed for older children (11+) but many of the pictures are for a younger audience than that. Also, the norms on which this assessment is based are still from Blank et al (1978) so they are really only appropriate for 3-6 year-olds.

<u>1.7.1.4 Language for Thinking. A structured approach for young children (Parsons & Branagan, 2005)</u>

Language for Thinking is a British book which includes assessment of and intervention for verbal reasoning based on the Blank model (1978). It was developed by two Speech and Language Therapists (SLTs) with extensive experience of working with language impaired children. Their research is at the least robust level of evidence-based practice, expert opinion (level iv) (Dollaghan, 2007).

The main focus is on improving verbal understanding**,** in particular verbal reasoning skills and inferential comprehension. It also includes strategies however for developing reading comprehension. It is designed so that SLTs and teachers can share targets and monitor progress.

There are three assessments of spoken and written language to be used in parallel. One assesses each child's starting level and two others monitor progress.

It is designed for typically developing children aged 4 to 7 years and for children with language and or learning difficulties aged 4 to 11+ years. Parsons and Branagan's prerequisites for the assessment and subsequent intervention programme are: understanding of Who/Where/What/Why question types; at least three key word level verbal comprehension; and the ability to speak in short phrases.

The assessments and interventions are based on Blank's levels two to four. These correspond to language levels A, B and C in *Language for Thinking*.

Level A covers early abstract understanding, with the child answering questions such as:

What's happening?

Who is in the picture?

Who went swimming?

What were Lisa and Jane talking about?

Tell me some things you could write with.

Level B moves on to manipulating what the child sees and hears.
Example questions are:

How do others feel?

What could others say?

What will happen next?

What is the cause of ....?

Tell me how the children planted the seeds

What is a dice?

What is something you shouldn't do with a book?

Level C moves up to analysing, explaining and reasoning. Example questions at this stage are:

Why will X happen?

Why shouldn't .....?

What would she do if.......?

How can you tell .....?

How do you know how old Caroline is?

What is the same about grandmothers and grandfathers?

If Jamal didn't have a tray what could he do?

The assessment is not a formal test and is not used to work out a developmental language age of the child. The score determines what level the child should start the intervention programme. Scoring guidelines and examples of different correct responses are provided on the same lines as Blank et al's PLAI-2 assessment (2003).

Each question can gain a score of 0-3 depending on the quality of the response. The authors acknowledge that this scoring system does rely on some subjective judgement. One of the assessments is provided in Appendix iii.
A positive feature of this assessment is that it is British so uses British vocabulary. The current researcher's criticisms of this model are the same as for the original Blank (1978) model. In addition it has not been standardised on any children, many 'Why' questions occur in Level B in this programme whereas in the Blank model they are only in level 4, and some questions are difficult to mark 0-3 (eg question 13 'What is John sitting on?').

Inferential comprehension tests not based on the 'Language of Learning' model

Standardised assessments

1.7.1.5 Inferential comprehension sub-test from the ACE - Assessment of Comprehension and Expression 6-11 (Adams et al, 2001)

This standardised sub-test from the British ACE assessment consists of a picture of a family and police in the kitchen of a house that has been burgled and a short verbal description. The child's ability to infer information from this scenario is assessed via nine verbal questions. These are:

1. Why was the dog barking?
2. Why is the policewoman there?
3. Why did the burglar break in at the **back** of the house?
4. Why do you think the burglar **only** took the watch?
5. What clues will the police find about who broke in? (Prompt: Can you think of any other clues?)
6. How does the family feel now? (Prompt: How else might they feel?).
7. Why would someone steal something? (Prompt: Can you think of any other reasons?)
8. What will the family do now because of the burglary? (Prompt: Can you think of anything else they will do?)
9. Should all theft be treated in the same way? (Prompt: Why do you think that?)

Specific scoring guidelines are given along with examples of acceptable and unacceptable responses for each question. A scoring table is given in the test manual, p37-38. For items 1-4 the scores are 0 for an unacceptable response and 1 for an acceptable response. Items 5-8 elicit more elaborate responses so are scored 0 for an unacceptable response, 1 for an acceptable response and 2 for two or more acceptable responses from their table. Item 9 gets a score of 2 if the child produces an acceptable response which includes justification and 0 if not. Some flexibility in answers is accepted if the overall meaning is maintained. The maximum possible score is 14. Standard scores are provided for the seven year-groups from 6:00 to 11:11 (130 per year group).

Positive features of this sub-test are that it is British so uses British vocabulary and it was standardised on a large population of children (790) aged 6:00 to 11:11 (130 per year group).

The current researcher's criticisms of this sub-test are: burglaries are not often within children's life experience; it is based on only nine questions; question 5 can be answered directly from the text and/or picture; some questions contain moral elements (eg questions 7 and 9); some of the responses the authors describe as unacceptable other markers would find acceptable (for example, Question 2: 'Why is the policewoman there?' Adams et al don't accept 'To see if she can do anything to help' or 'She's going to catch the burglar').

Validity and reliability data are given for this ACE sub-test. They will be discussed in comparison to the new test's findings in the discussion chapter.

1.7.1.6 The Listening Test (Barrett et al, 1992)

The Listening Test is an American test appropriate for children aged 6 to 11 years. It takes 35 minutes to complete and needs to be administered by a trained professional (eg Speech and Language Therapist, psychologist, teacher etc). There are no basal and ceiling scores. Each sub-test is carried out in its entirety.

The test consists of 5 sections: A. Main Idea, B. Details, C. Concepts, D. Reasoning, E. Story Comprehension. Each section contains 15 questions and each question scores a maximum of 1. An example question from the most relevant sections (A, B and D) is given below. Examples from sections C and E are given in Appendix iii.

A. Main Idea

A scenario is read out:

Ray's friends came to his house. They brought presents for Ray and played games. Ray blew out all the candles on the cake.

Question (Q)1: What am I talking about?

Answer (A): Birthday

B. Details

1. Tracy and her dad took off the training wheels. Tracy wobbled a little at first, but then she started riding easily.

Q. What was Tracy riding?

A. Bike

D. Reasoning

1. Ms Burns has a surprise for her class inside a box. There are small holes in the box. Something inside the box is moving and purring.

Q. What do you think is inside the box?

A. Cat/kitten

Age equivalents are provided for the raw scores for varying age bands ranging from 5:06 to 11:07.

The current researcher's criticisms of this test are that no repetitions of any of the questions are allowed which places a huge demand on auditory memory for the child. There is considerable overlap between each section so it is not clear what aspect each section assesses. Some questions depend on specific vocabulary knowledge rather than purely reasoning ability. Finally, it contains many American vocabulary items and concepts that do not transfer easily to the UK, for example: janitor, root beer, nickel, vacation, yard, social studies contest, chalkboard, bad grades, touch football, garbage can, mailbox, the mall, bus schedule, first grade, training wheels, baseball, softball, cookie.

An updated version of the Listening Test was developed by one of the original authors and two other researchers in 2006: The Listening Comprehension Test 2 (Huisingh et al, 2006). It covers the same age range, 6-11 years, and it assesses how children attend to, process, and extract meaning from verbal language. The scenarios used are described as 'real-life' classroom listening situations. It differs from the original *Listening Test* in that each passage has four question types: Main idea, Details, Reasoning and Vocabulary. An example is given below. Passage 1 is the same as the first passage in the original *Listening Test* but the other passages are not the same. An example of a different one (passage 3) is given in Appendix iii. There is also a separate fifth sub-section 'Understanding Messages'.

Passage 1 (see A: Main Idea above)

A. Main Idea

Q1. What am I talking about?

Only acceptable response any reference to birthday/party

B. Details

Q2. Where was the party?

Only acceptable response: any reference to Ray's/his house

C. Reasoning

3. What might Ray's friends have done to get ready for his party?

Only acceptable responses: any reference to get/buy/wrap present/gift, get a card, get cleaned up/dressed, get a ride/ how they get to the party, blow up balloons, decorate house, get/decorate cake

D. Vocabulary

4. What is another word for *gift*?

Only acceptable response: present

E. Understanding Messages is the last subtest on the assessment. It requires the child to interpret 'real-life' messages such as public announcements. An example is given below:

> Message:
>
> Parent conferences are next week, so school will dismiss 30 minutes early on Tuesday, Wednesday, and Thursday.
>
> Question: What days are parent conferences?
>
> Question: Why will students be dismissed early next week?

Positive features of this test are that it was standardised on a large normative sample of 1,504 children (215 per year group), the participants were randomly selected and they reflected the American national school population demographics for race, gender, age, and educational placement. In addition, the normative data had high reliability and validity, and raw scores, age equivalents, standard scores and percentile ranks were provided.

Criticisms of the test are the same as for *The Listening Test* and some passages require knowledge of American educational procedures and American history (eg dismissing students, the Pilgrim Fathers).

1.7.1.7 Test of Language Competence – Expanded Edition (Wiig and Secord, 1989)

This American test has two levels. Level 1 covers the age range 5-9:11 and level 2 the age range 9-18:11. It measures higher-level language functions and takes under an hour to administer. There are 4 sub-tests: 1. Ambiguous sentences, 2. Listening comprehension, Making Inferences, 3. Oral expression - Recreating Speech Acts, and 4. Figurative Language. Level 2 also has a supplemental subtest: Remembering word pairs. Raw scores are converted to percentile scores.

Dennis and Barnes (2001) describe the making inferences sub-test as a 'script inferencing task' in which the child has to make logical inferences from scenarios based on knowledge of possible causal links. The scenarios involve common experiences from home, school and other familiar settings.

Examples from the second sub-test 'Making Inferences' for levels 1 and 2 are given below. The child has to choose the two correct answers from a choice of four. Level 1 has accompanying pictures.

Level 1 (16 scenarios)

1. It was hot, so Kim went to the swimming pool. He was mad when he couldn't swim.

Kim couldn't swim because:

a. The lifeguard was watching the kids (No)

b. His friends were at the pool? (No)

c. The pool was being cleaned? (Yes)

d. He left his swimsuit at home? (Yes)

Level 2 (12 scenarios)

1. Jack went to a Mexican restaurant. He left without giving a tip

Jack didn't leave a tip because:

a. The restaurant closed when he arrived? (No)

b. He only had enough money to pay for the meal? (Yes)

c. The food and the service were excellent? (No)

d. He was dissatisfied with the service? (Yes)

Nippold (1998) rates this as a good assessment that includes subtests for understanding linguistic ambiguity and figurative expressions.

Positive features of this test are that it gives standard scores for age groups ranging from 5:00 to 18:11 and it has strong reliability scores (Level 1 is .86 to .92 and Level 2 is .75 to .82).

The disadvantage for a British population is that much of the vocabulary is American. In addition, in this sub-test, the child has a 25% chance of selecting the correct answer by guessing from the 4 choices given.

However, the format of this sub-test, testing understanding of inferential comprehension in a short scenario, is a useful one as it assesses above sentence level inferential comprehension. It therefore formed the basis for the current researcher's 'making inferences from short passages' sub-section of her new assessment.

1.7.1.8 Test of Problem Solving (TOPS) (Zachman et al, 1984)

This American assessment covers the age range 6:00-11:11. The authors describe it as an *expressive test* that assesses a child's reasoning and thinking skills in real life contexts. It consists of 15 line drawings and 50 questions associated with these. A verbal response from the child is required.

It is composed of 5 sub-tests: 1. Explaining inferences, 2. Determining causes, 3. Negative why questions, 4. Determining solutions, Avoiding problems. There are no basal or ceiling cut-off points. The test is administered in its entirety. The child's responses are given a score of 2, 1 or 0 points depending on the appropriateness of the content and syntax. A child whose scores fall significantly below the norms is identified as having delayed verbal reasoning skills.

Each picture has questions from each of the 5 thinking tasks, eg

Picture 1: Two people dining in a restaurant

Question (Q) 1. How do we know these two people are at a restaurant? (Explaining an inference)

Q2. Why did they decide to go to a restaurant? (Determining causes)

Q3. Why won't they wash the dishes after they eat? (Negative Why question)

Q4. The waitress brought them hamburgers and French fries, but they ordered spaghetti. What could they do? (Determining solutions)

Q5. What could the waitress have done to keep from making this mistake? (Avoiding problems)
Sub-tests 1 and 2 both assess inferencing ability. Further examples of inference questions are given below:
<u>Picture 2: Mum trying to talk on phone with boy playing loud music</u>
Q1. How do we know Mum is having trouble talking on the phone?
Q2. Why does the boy have the music turned up so loudly?
<u>Picture 3: American football game</u>
Q1. One of the teams just scored the winning point. How can we tell from this picture?
Q2. One player is being carried around the field on his teammate's shoulders. Why are they doing that?

The positive features of this test are that the age range is similar to the researcher's new assessment, it was standardised on 842 children (140 per year group), and age equivalents, standard scores and percentile ranks are provided for each sub-test.
The current researcher's criticisms of the test are that it is described as an expressive test when it is primarily assessing verbal comprehension; the scoring system includes syntax so disadvantages children with expressive language impairments; the 15 scenarios are described by the authors as being socially and conversationally relevant and commonly experienced by young children, but some are outdated or specific to American culture (eg American football); and each scenario has a different number of questions and question types associated with it, ranging from 2 to 5 questions per picture, with the question weightings not being justified by the author. In addition, the scoring guidelines are very ambiguous. The current researcher would give full marks to many of the test developers' zero scores (for example for Picture 10 depicting a boy with a bike with a flat tyre: Q1. 'How did the boy's bike get a flat tire?' (Determining causes) I would accept their 0 scores of: 'He ran over a pin/tack, thorn/rock' or 'He ran over a sharp rock').
A confusing aspect is that the test is designed for ages 6:00 to 11:11 but age equivalents are provided for 3:07 to 15:11 for question types 1 and 2.
A study by Bernhardt (1990) questions the content validity of the response scoring criteria of the TOPS. She states that the scoring system relies on clinical judgement,

which can be very subjective and that no reliability (intra- or inter-examiner) or validity data are provided in the test manual She also points out that the pictures and accompanying questions are not arranged in developmental order. She tested the content validity of the TOPS by attempting to replicate the scores of 50 examples given in the manual using 10 Masters level Speech and Language Pathology students. All of the students said the scoring was difficult and the manual guidelines were ambiguous. The mean agreement with the manual values was only 48.4%. Their findings suggest the TOPS scoring system is seriously flawed, so using the normative data from the test is not appropriate or clinically useful.

Bowers et al (1994) developed a revised Test of Problem Solving – Elementary (TOPS-E), age range 6-11. It is in the same format as the original TOPS but uses photographs rather than line drawings.

Zachman et al (1991) also developed the Adolescent Test of Problem Solving for 12 to 18 year-olds and a revised edition was brought out 16 years later (Bowers et al, 2007). The five critical thinking areas targeted in this assessment are: Making Inferences, Determining Solutions, Problem Solving, Interpreting Perspectives, and Transferring Insights. Both revised editions provide standard scores, percentile ranks and age equivalents. They also include strong reliability and validity scores.

The current researcher's criticisms of the Adolescent TOPS assessments are that they involve a lot of general and curriculum-specific knowledge (science, history, environmental studies, etc). Some questions also involve moral and value judgements.

1.7.1.9 Test of Auditory Reasoning and Processing Skills (TARPS) (Gardner, 1993)

This American assessment is designed for 5-13:11 year-olds. It aims to measure children's verbal reasoning and thinking skills. It takes 10-20 minutes to administer and can be scored in 10 minutes. It assesses the following areas of auditory reasoning and processing:

1. General Information
2. Arithmetic Reasoning
3. Verbal Absurdities
4. Finding Reasons
5. Analogical Completion

6. Comprehension
7. Directional Orientation
8. Similarities

The questions are arranged progressively according to difficulty. The test is administered and scored in its entirety so no sub-test age equivalent scores are given. Some examples of 'Why Questions' in the assessment are:

2. 'Why do houses have doors?' (Acceptable Answers (AAs) – Get in/or out).
11. 'Why do houses have chimneys?' (AAs – Let smoke/exhaust/fumes out).
22. 'Why do swimmers use flippers?' (AAs – Swim faster/better/more power).
40. 'Why is it better to send a message by fax than by mail?' (AAs – faster/more convenient).
43. 'Why does a cork float on water?' (AAs – Lighter/Water is heavier/More buoyant/Not as heavy as water).
46. 'Why is it better to pay by credit card than by cash?' (AAs – Don't need cash/ Don't have to pay for 30 days/Safer/A record of purchases).
47. 'Give 2 reasons why people should not take drugs' (AAs – Expensive/ inability to function/illegal/ruin your health/effects on others/dangerous/bad for you/ transmission of aids and other diseases).

A positive feature of this test is that it was standardised on a sample of 1140 children aged 5-13:11 (126 per year group).

The current researcher's criticisms of this assessment are that despite the author claiming the test tries to be as culture-free or cross-cultural as possible, many vocabulary items and concepts apply only to American children; many questions are American general knowledge and maths questions; and some questions are outdated (Q40) or incorrect (Q46 – there may also be disadvantages to using a credit card, ie an extra charge).

1.7.1.10 Comprehensive Assessment of Spoken Language (CASL) (Carrow-Woolfolk, 1999)

This is an American Assessment covering the age range 3 to 21 years. The author states that it provides an in-depth, research-based test of oral language skills and that it is ideal for assessing verbal language delays and disorders, dyslexia and aphasia.

It consists of 15 individually administered stand-alone sub-tests with a multiple choice answer format. No literacy skills are required to complete the test.

The sub-tests are grouped under four areas:

1. Lexical/Semantic**:** Comprehension of basic concepts, antonyms, synonyms, sentence completion and idiomatic language.
2. Syntactic**:** Syntax construction, paragraph comprehension, grammatical morphemes, sentence comprehension and grammaticality judgment.
3. Supralinguistic**:** Nonliteral language, meaning from context, inference and ambiguous sentences.
4. Pragmatic**:** Pragmatic judgement (knowledge and use of appropriate language).

Examples from the most relevant sub-sections for inferential comprehension are:

Nonliteral language:

7. Mom came into Trevor's room and asked 'When did the tornado come through?' The target response is that the room was very messy.

12. Mom looked in Marshall's ears and said 'You could grow flowers in there'. The target response is that his ears were dirty.

Drawing inferences:

19. 'The builder had to change the size of the doorways in the basketball player's house. Why?' The target response is that the player was too tall.

22. 'Dad could not find the candles, so Chuck could not do his homework'. Why? The target response is that the power was out.

The assessment was standardised on a sample of 1,700 participants. Norms are provided at 6 month intervals for the 3-4 year-olds, 1 year intervals for the school-aged children and multi-year intervals for the older age groups.

The test provides standard scores, percentiles and age equivalent scores.

The author reports strong internal reliability scores (0.64-0.94) and test-retest reliability scores.

Brandel et al (2011) assessed 216 children aged 6-8 years with specific language impairment (SLI) on the Comprehensive Assessment of Spoken Language (CASL) (Carrow-Woolfolk, 1999) and the Test of Language Development – Primary, Third edition (TOLD-P:3, Newcomer and Hammill, 1997). They were testing the concurrent and construct validity of these oral language assessments to determine the strength of association between their language scores. They hypothesized that the overall language performance of the children with SLI should be significantly and positively correlated on both tests, as they both assess decontextualized verbal language and are norm-referenced. They found that the validity values they measured differed from those cited in the test manual. They did find a significant overlap in the results of each test but also found substantial differences. They concluded that the assessments measured different subsets of language abilities. Their concern was that the tests could provide contradictory findings when used with the same child. They concluded that, when assessing oral language abilities the tester should not rely on just one formal assessment and should collect language data on the child from a variety of sources.

1.7.1.11 Mount Wilga High Level Language Test (Christie et al, 1986) and the revised version (Simpson, 2006)

The original assessment was devised by Christie et al (1986), the Speech Pathology team at Mount Wilga Rehabilitation Centre, Australia. It was a screening assessment used to assess a predominantly head-injured caseload who presented with mild language problems that significantly affected their functional communication. Available standardised tests for aphasia were of limited use for these clients as they did not pick up their subtle high-level language comprehension difficulties. The test was administered to 100 typically developing participants aged 15 to 25 years and some norms obtained.

In 2006 a British speech and language therapist, Simpson (2006), updated the assessment with the original authors' permission and adapted it for the UK.

There are 8 main test sections with sub-sections within them. These are: naming skills, verbal explanation, planning, auditory memory, auditory comprehension, reading comprehension, written expression and numeracy. Examples of these are given in Appendix iii. The most relevant examples for an assessment of inferential comprehension from sections 2 and 6 are given below.

*2G. Absurdities***:** What is ridiculous about these stories?

1. An old lady said 'I'm no longer able to manage my usual walk around the block each day. Now I can only go half way round and back again that's all'.

2I. Verbal reasoning:

2. Give me two reasons why people like to eat in restaurants.

6S. Inferential Paragraphs:

Circle the correct answer:

1. Most men would be insulted if they were asked to earn their wages by throwing stones over a wall and then throwing them back again. Men generally like to work at jobs that they think are....

a) meaningful b) uncertain c) underpaid d) tiring

This is a test designed for adults not children, standardised on a young adult population, so is not directly applicable to an assessment of children's inferential comprehension. However, the newly created inferential and idiom comprehension test could also prove to be a useful tool for differential diagnosis of some acquired neurological disorders so it is worth examining the adult assessments that include inferential comprehension.

The current researcher's criticisms of this test are: the normative data are very limited (only 10 per year group); some questions are outdated and no longer culturally acceptable (eg the inferential paragraph 1 should say 'people' not 'men'). The original authors describe the test as being useful for providing a baseline and a framework for management but acknowledge that the test has no cohesive theoretical basis and is very difficult to standardise or validate. This could only be done on a sub-test by sub-test basis.

<u>Other assessments which include inferential comprehension questions</u>

The Clinical Evaluation of Language Fundamentals (CELF) 4 UK (Semel et al, 2006), the New Reynell Developmental Language Scales (NRDLS) (Edwards et al, 2011), and the Expression, Reception and Recall of Narrative Instrument (ERRNI, Bishop, 2004) include a few questions that assess inferential comprehension. These are in the Understanding Spoken Paragraphs sub-section in the CELF-R UK, the Inferencing section of the RDLS Section H and following the stories in the ERRNI. However, it is not possible to extract a separate score for inferential comprehension from any of these assessments.

In England, pupils sit the verbal subtest of the Cognitive Abilities Test (Strand 2006, in Spencer et al, 2016), when they start secondary school. This tests verbal reasoning and is used to predict educational progress.

Non-standardised Assessments

These non-standardised assessments have been created by expert practitioners to assess areas of specific difficulty in their client groups that could not be tested by existing formal assessments.

1.7.1.12 Black Sheep Press: Informal assessment of social language and communication skills for children in primary school (Ross, 2011)

This is an informal assessment compiled by Ross and her SLT colleagues in the Dumfries and Galloway paediatric SLT service. It aims to help in the differential diagnosis of Autistic Spectrum Disorders. It uses many of the Black Sheep resources as part of the assessment.

The following areas of communication are informally tested: recognising, labelling and demonstrating emotions; describing situations that cause different emotions; matching emotions to scenarios presented orally or in picture clues; self-assessment of social situations; social understanding (understanding idioms in context, describing feelings and verbal reasoning; talking about friends and school); school based understanding; understanding of idioms out of context; and observations of interaction skills).

The verbal reasoning questions in section 5 (social understanding) are based on picture scenarios. Two examples of these are:

1) Picture of a teacher tending to a boy who has fallen over and hurt his knee whilst a girl is tapping her on her shoulder:

Verbal reasoning questions:

What will make Luke feel better?

Mrs Lawson is too busy to listen to Jenny so what can Jenny do instead?

2) Picture of a boy who has kicked a ball at a window and smashed it.

Verbal reasoning questions:

Why is it important to sweep up all the glass?

Where is it safe to play football?

Where is it not safe to play football?

1.7.1.13 Canterbury and Thanet Verbal Reasoning Assessment (Johnson, 1998)

This informal assessment of verbal reasoning is made up of examples from various IQ tests, in particular the Stanford-Binet Intelligence Scales, (Terman & Merrill, 1960). The sub-tests in Johnson's assessment are: association and explanation, cause-effect, opposite analogy, auditory memory, definitions, analogy, word recall, similarities and differences, inference, and deductions and absurdities. The author reports that the 'sequential logic' items (cause-effect, inference, deduction, and verbal absurdities) provide the best indication of the child's overall cognitive ability. Age levels by which typically-developing children should be able to answer the questions are given for the various question types. A few examples of the sequential logic questions are given below and more are given in Appendix iii:

By 3:06 to 4 years

Cause and Effect:

What must you do when you are hungry?

What must you do when you are sleepy?

What must you do when you are cold?

By 4 years

Explanation:

Why do we have houses?

Why do we have books?

By 4:06 to 5 years

Cause-effect:

What must you do when you have lost something?

What must you do before you cross a road?

By 8 years

Inference:

a) Julie heard a big 'bang' and came running outside. There were nails all over the road and a car had just stopped beside the road. What was the 'bang'?

b) A man and a lady were sitting in a restaurant. They had just eaten a big dinner. The waiter brought the bill. The man looked at it, and then seemed worried and embarrassed. Why?

By 13 years

Inference:

c) An Indian who had come to town for the first time in his life, saw a white boy riding along the street. As the white boy rode by, the Indian said, ‘The white boy is lazy; he walks sitting down’. What was the boy riding on, that made the Indian say ‘he walks sitting down.’?

The current researcher’s criticism of this assessment is that the main source it is drawn from is very outdated and no longer culturally acceptable. The original Stanford-Binet scale was developed in the 1920s.

<u>1.7.1.14 High level language and communication screening test (McCartney, 2005)</u>

This is a non-standardised British screening assessment suitable for 4-10 year-old children. The author gives rough age ranges based on existing findings from the literature (but she does not state what these are).

There are four sections to the assessment: 1. Social Communication Skills observation areas, 2. High Level Comprehension, 3. Auditory processing and Memory, 4. High Level Semantics.

Some examples of questions from section two are given below:

<u>1. Understands why-because reasoning (Appropriate age range 3:6-4:6)</u>

Q. Why do you need a towel when you go swimming?

<u>2. Understands how reasoning (Appropriate age range 4:0-4:6)</u>

Q. How do you know when you need to wear glasses?

<u>3. Understands time concepts (Age range 4:06-5:06)</u>

Q. When do trees lose their leaves?

<u>4. Can verbally sequence events over 4-5 steps in an appropriate order (Age range 5:6-6:0)</u>

Q. Tell me how to make a sandwich?

<u>5. Can predict what would happen if...? (Age range 7-8)</u>

Q. What could happen if you do not stop at traffic lights?

<u>6. Can generate a personal opinion (Age range 7-8)</u>

Q. What would you do if you were putting your sock on and found a hole in it?

<u>7. Can predict an outcome from a given scenario (Age range 7-8)</u>

Q. Mrs Jones returns from the shops to find water all over her kitchen floor. She had left the washing machine on to finish its cycle. What will happen?

8. Can draw inferences from given situations (Age range 8:0-8:06).

a) Sam was upstairs in bed when he heard the postman push some letters through the letterbox. He also heard Roxy the dog barking loudly. When he got downstairs there were no letters on the floor. What could have happened?

The current researcher's criticisms of this assessment are that the extracted norms are not referenced and the 'how' reasoning questions range from simple 'how much?' or 'how many?', which are dependent on a number concept, to the more complex 'how do you know?', which involves theory of mind

1.7.1.15 The Right Hemisphere Language Battery (Bryan, 1995)

The Right Hemisphere Language Battery (RHLB) was devised by speech and language therapist Karen Bryan to provide quantitative and qualitative assessment of the communication impairments in (predominantly acquired) right-hemisphere neurological impairments. Subtests focus on right-hemisphere skills such as lexical-semantic processing, high level language processing and prosody. It consists of six sub-tests: two tests of metaphorical language (one with pictures and one written), one of inferential comprehension, one of comprehension of humorous material, one of lexical semantics, and one of the appropriate use of emphatic stress. It also includes a 5-point rating scale for discourse analysis which provides a subjective score for appropriateness of expressive language and turn taking ability. The battery was validated on 30 patients with right-hemisphere damage of vascular origin and 10 non-vascular patients.

Thomson et al (1997) questioned the clinical usefulness of the RHLB for assessing communication impairments in patients with tumour rather than vascular right-hemisphere damage. They used it with 20 people with a right-hemisphere tumour and 20 neurologically typical controls. Their results demonstrated that several people in both groups performed very poorly in the metaphor, inference and humour tests. Three of the controls only scored 7/12 in the inference section and only 1 control participant obtained the maximum score of 12. The subjects also said the inference passages were too long and contained too much information, thus making it more of a memory test.

Talarowska et al (2012) on the other hand, in a single case study, found the Polish version of the RHLB to be a useful tool in the assessment of the communication skills of a 29 year-old Polish adult with schizophrenia. The patient demonstrated difficulties similar to those of patients with right-hemisphere impairments. He had difficulty interpreting emotional intonation and understanding complex language and humour.

Table 1.2 below summarises the advantages and disadvantages for British children of each of the above inferential comprehension assessments.

Table 1.9 The advantages and disadvantages of current inferential comprehension assessments

Test Name	**Advantages**	**Disadvantages**
Blank Language of Learning Model and the PLAI	1.Standardised published test 2.Inter-rater and test-retest reliability within acceptable range for quantitative scoring.	1.Small standardisation sample (20 per year group) 2.American 3.Subjective qualitative scoring 4.Levels 3 and 4 not developmental
Liverpool Language Screen	1.Standardised 2.British 3.Reasonable sample (50 per year group)	1.Unpublished 2.No supporting inferential statistics
TALC1 and TALC2	1.British	1.Not standardised 2.Based on American data for 3-6 year-olds 3.No marking criteria provided
Language for Thinking	1.British	1.Not standardised
Inferential	1.British	1.Limited

Comprehension Sub-Test of ACE	2.Standardised 3.Large sample (130 per year group).	assessment of inferential comprehension (only 9 questions based on 1 composite picture). 2.Some ambiguous marking criteria
The Listening Test	1.Standardised 2.Very large standardisation sample (1,504 children, 215 per year group). 3.High reliability and validity	1.American 2.No repetitions of test questions allowed (increasing the load on auditory memory).
Test of Language Competence, Sub-test 2: Making Inferences	1.Standardised 2.High reliability	1.American. 2.Choose from 1/ 4 pictures (so 25% chance of guessing correct answer)
Test of Problem Solving (TOPS)	1.Standardised 2.Large standardisation sample (842 children, 140 per year group)	1.American 2.Is described as an expressive test and the scoring system includes syntax. 3.Ambiguous scoring guidelines
Test of Auditory Reasoning and Processing Skills (TARPS)	1.Standardised 2.Large standardisation sample (1140, 126	1.American 2.Outdated content

	per year group)	
Comprehensive Assessment of Spoken Language (CASL)	1.Standardised 2.Very large standardisation sample (1,700, ca 212 per year group) 3.High internal reliability	1.American 2.Concurrent and construct validity in manual not upheld by other researchers
Mount Wilga	1.British version 2.Standardised	1.Adult assessment 2.Very small standardisation sample (100 people aged 15-25, 10 per year group) 3.Some outdated content 4.No cohesive theoretical basis
Informal assessment of social language and communication skills for children in primary school	1.British	1.Not standardised
The Canterbury and Thanet Verbal Reasoning Assessment	Multiple origins	1.Not standardised 2.Based on very outdated non-British assessment (from 1920s)
High level language and communication screening test	1.British	1.Not standardised 2.No theoretical basis given
The Right	1.British	1.Adult assessment

Hemisphere Language Battery		2.Not standardised 3.Subjective scoring system

<u>1.7.2 Assessments of idioms and other non-literal understanding</u>

<u>Standardised assessments</u>

<u>1.7.2.1 ACE - Assessment of Comprehension and Expression 6-11 (Adams et al, 2001) – Non-literal comprehension sub-test</u>

This is one of the extension sub-tests within the ACE assessment. It assesses the understanding of non-literal phrases in context. The sub-test has two parts. In the first section the child has to choose the picture that represents the correct interpretation of the idiom (out of a choice of four). In the second section the child has to choose the correct interpretation of idiomatic phrases from a choice of four written alternatives (which are also read out by the tester). The idiomatic phrases are given in italics below.

<u>Section 1</u>

1. The cheese was mouldy so my mum told me to *throw it away.*
2. There was a war in outer space. The aliens *blew up* one of our space ships.
3. Our teacher was really angry because we spent the whole lesson *messing around.*
4. Kim and Neeta had a big argument. Kim said she was sorry and they *made up*
5. We can't go to the park now, it's started *bucketing down.*
6. We were going to the cinema when Mum's mobile *went off* in her bag.
7. I parked my car outside but somebody's *boxed me in* – I can't get out.

Section 2 assesses the following non-literal phrases: pick something up, get on, call something off, die down, cave in, face up to something, cross something out, get over something.

The current researcher's criticism of this sub-test is that some of the assessed terms are idioms and some are literal phrasal verbs (eg throw it away, blow up).

Validity and reliability data are given for this ACE sub-test. They will be discussed in comparison to the new test findings in the discussion chapter.

1.7.2.2 Test of Language Competence (Wiig and Secord, 1989)

Sub-test 4 in this test as a test of idiom comprehension. Dennis and Barnes (2001) describe the idioms in this subtest as structural, orientational, ontological, or part/whole. Children hear a context-providing introductory sentence and then an idiom. In level one they have to define the idiom. In level two they have to select the correct, figurative meaning from a choice of four with incorrect choices that include a matching interpretation, a literal interpretation, an opposite interpretation, and a non-related interpretation. Raw scores are converted to percentile ranks.

Level one, 5:00 to 9:11, Sub-test 4, tests the following idioms in context. The child has to explain what each idiom means.

For example:

1. This is what a teacher said to a boy:

   'Let me give you a hand.'

2. This is what a girl said about her mother:

   'Mom looked really low today.'

3. This is what a girl said about a project:

   'I want to go it alone.

4. This is what a boy said about a birthday party:

   'We are having a ball.'

The other idioms tested at this level are given in Appendix iii.

Level two, 9:00 to 18:11, is a multiple choice test of idiom comprehension in context. The child chooses between a nonrelated (N), literal (L), opposite (O) or matching (M) response. For example:

1. Situation: A boy talking about a girl at a school dance.

   Expression: 'She sure casts a spell over me.'

   a. In her life, every day is Halloween (N).

   b. She spells much better than I (L).

   c. I am out from under her spell (O).

   d. She is totally bewitching to me (M).

The other idioms assessed at this level are given in Appendix iii.

The current researcher's criticisms of this sub-test are that: many American idioms are not used in British English, a definition task (used in level 1) is more difficult than a multiple choice task (used in level 2), and many of the correct responses to an idiom example in level 2 are other idioms, examples of which are given in Table 1.3 below:

Table 1.10 Idioms defined by idioms in the Test of Language Competence

Idiom	Accepted definition
rough sailing	facing a hard road
high on the totem pole	top dog
stew over it	toss it around more
I can't swallow that	that smells really fishy
up in the air	up for grabs
all behind us	it's water under the bridge

Rinaldi's (1996) criticism is that this assessment does not explore the role of context in the comprehension of these idioms. However Wiig and Secord (1989) claim that their introductory sentence does provide some context.

1.7.2.3 The Fullerton Language test for Adolescents (Thorum, 1986)

This language assessment is designed for 11 to 18 year-olds. It was standardised on 762 students. Sub-section eight of this assessment is idiom definition. Twenty idioms are assessed. These include 'pull the wool over his eyes', 'see eye to eye' and 'the cat's got your tongue'.

Nippold (1998) is critical of this assessment, claiming that it includes items that are too easy and Rinaldi (1996) points out that the idioms are not assessed in context. The test is now out of print.

1.7.2.4 Test of Word Knowledge (Wiig and Secord, 1991)

This assessment has a figurative usage sub-test aimed at 8-17 year-olds. It is a multiple choice test. The first question asks the child to identify what an idiom means out of a choice of four. The second question asks the child to select the idiom that matches the literal language, also out of a choice of four. There are 42 questions in total

For example (the correct answer is given in italics below):

1. What does ‘cut that out’ mean?
    a. Give an answer
    b. Go slowly
    c. Run away
    d. Stop it
2. Which one tells about someone who is not being noisy?
    a. Busy as a bee
    b. Quiet as a mouse
    c. Sly as a fox
    d. Eats like a bird

The other idioms covered in this sub-test are given in Appendix iii.

Nippold (1998) evaluates this as a good assessment which examines sophisticated aspects of language such as knowledge and use of abstract words, the ability to define words, and the understanding of adverbial conjuncts.

The current researcher’s criticisms of this sub-test are that the idioms are American and it uses phrases and compound words as well as idioms.

Informal assessments

1.7.2.5 The Idiom Comprehension Test (ICT) and the Korean Autism Social Language Test (KASLAT)

Qualls and Harris (1999) developed The Idiom Comprehension Test (ICT) for their exploration of the effects of familiarity on idiom comprehension in 10 year-old American children. The ICT contains 24 idioms: 8 of high familiarity, 8 of moderate familiarity and 8 of low familiarity.

The high familiarity idioms are: let off steam, go around in circles, put one’s foot down, breathe down one’s neck, read between the lines, put their heads together, skate on thin ice, beat around the bush.

The moderate familiarity idioms are: go into one's shell, strike the right note, keep up one's end, cross swords with someone, go against the grain, blow the cobwebs away, make one's hair curl, throw to the wolves.

The low familiarity idioms are: take down a peg, vote with one's feet, paper over the cracks, hoe one's own row, talk through one's hat, lead with one's chin, rise to the bait, have a hollow ring.

Lee et al (2015) designed the Korean Autism Social Language Test (KASLAT) to test idiom comprehension of children with high functioning autism spectrum disorder in Korea.

However, the samples were very small in both of the above studies and the idioms were American or Korean. In addition, the ICT used low, moderate and high familiarity idioms whereas the current researcher deliberately uses only high frequency idioms in the HICIT.

1.7.2.6 Black Sheep Press: Informal assessment of social language and communication skills for children in primary school (Ross, 2011)

This informal screen assesses comprehension of idioms both in and out of context. Section five tests understanding of idioms in context: Two examples of these are given below (the initial scenarios are the same as those given for the verbal reasoning questions in section A12 above but they are included again here to show what contexts are provided for the idioms):

1) Picture of a teacher tending to a boy who has fallen over an hurt his knee whilst a girl is tapping her on her shoulder:

    Jenny is tapping Mrs Lawson on the shoulder. She wants to speak to her but Mrs Lawson says 'Just a moment Jenny. I'm a bit tied up at the moment.' What does Mrs Lawson really mean?

    Luke is really making a lot of noise and fuss and he's only grazed his knee. Jenny says 'Oh, put a sock in it!' What does Jenny mean?

    Luke realises he's been making too much fuss. He says 'I'd better pull myself together.' What does Luke mean?

2) Picture of a boy who has kicked a ball at a window and smashed it.

Bill and David are playing football in the playground. Bill kicks the ball too hard and breaks the classroom window. Bill says 'Oh no, the teacher will hit the roof'. What does Bill mean?

David says he is 'shaking in his shoes.' What does David mean?

Bill says 'Come on – let's go and face the music'. What does Bill mean?

Section seven assesses understanding idioms out of context. The pictures are still presented however. The child is asked 'Tell me what this means':

Give me a hand, over the hill, lose your head, don't let the cat out of the bag, in hot water, she's cracking up, in over your head, shed some light on it.

1.7.2.7 Mount Wilga High Level Language Test (Christie et al, 1986) and the revised version (Simpson, 2006)

Sub-section 2H in this informal assessment is idioms. The adult has to explain what the following expressions mean: butterflies in your stomach; turn over a new leaf; fly off the handle; hit the nail on the head.

1.7.2.8 Understanding Ambiguity: An Assessment of Pragmatic Meaning Comprehension (Rinaldi, 1996)

This was a British assessment designed to assess a child's pragmatic skills. It was trialled on 139 children aged 6-14 years. It consists of two sections, Multiple Meanings in Context (MMC) and Inconsistent Messages of Emotion (IME).

The MMC sub-section contains homonyms (single words which have 2 or more different meanings, idioms (multi-word expressions whose meaning cannot be worked out from the individual words), and multiple meaning phrases (two word combinations occurring within the same phrase).

The following phrases, homonyms and idioms (in italics below) are assessed:

1. (phrase) Joe couldn't go to his sister's birthday party because he was *tied up*
2. (homonym) The road was *jammed* solid this morning.
3. (homonym) I've been getting very *short* with her.
4. (phrase) Her room is a real *pig sty.*
5. (idiom) He's *full of beans* today
6. (phrase) Have you two *fallen out* with each other?
7. (idiom) You need to *pull your socks up*

8. (idiom) He *drove me round the bend* this morning.
9. (homonym) The robber was *caught red-handed.*
10. (idiom) They are very *thin on the ground.*
11. (phrase) Don't *get carried away.*
12. (phrase) He was *beside himself.*
13. (idiom) He *got out of the wrong side of bed* this morning.
14. (homonym) That test really *threw* him.

The limitations of this test are that it was trialled on a very small number of typically developing children (15 per year group), it is now out of print and the current researcher disagrees with some of Rinaldi's terms (eg question nine would be better classified as an idiom than as a homonym). Rinaldi (1996) acknowledges that the sample size is not large enough to give normative scores and she describes her results as comparative data.

<u>1.7.2.9 The Children's Communication Checklist–Second Edition (CCC2) (Bishop, 2003)</u>

This informal assessment tests inferential comprehension and non-literal comprehension. It is a checklist that Dorothy Bishop developed to give a profile of children's communication strengths and difficulties. It contains a series of statements describing how children communicate in everyday situations which is completed by the child's speech and language therapist or teacher.

It uses a coding system for the frequency of behaviours observed:

0 for less than once a week or never

1 for at least once a week but not every day

2 for once or twice a day

3 for several times (more than twice) a day (or always)

Botting and Adams (2005) report that the scale was designed to distinguish children with pragmatic impairments from other children with more typical language impairments. Nine scales are included comprising speech, syntax, social relationships, interests, inappropriate initiation, making sense in conversation, stereotyped conversation, context and rapport. The last five of these make up a pragmatic composite scale that has helped in discriminating groups with PLI, ASD and SLI in the research. A threshold of 132 indicates PLI.

The main questions in this checklist that address non-literal comprehension are:

15. Misses the point of jokes and puns (though may be amused by nonverbal humour such as slapstick)

19. Gets confused when a word is used with a different meaning from usual: eg might fail to understand if an unfriendly person was described as ‘cold’ (and would assume they were shivering)

39. Fails to recognise when other people are upset or angry.

41. Is over-literal, sometimes with (unintentionally) humorous results. eg, a child who was asked ‘Do you find it hard to get up in the morning?’ replied ‘No’. You just put one leg out of bed and then the other and stand up” Another child who was told “watch your hands” when using scissors, proceeded to stare at his fingers.

54. Appreciates the humour expressed by irony. Would be amused rather than confused if someone said ‘isn’t it a lovely day!’ when it is pouring with rain.

57. Shows concern when other people are upset.

60. Realises the need to be polite – would pretend to be pleased if given a present s/he did not really like, and would avoid making personal comments about strangers.

Volden and Phillips (2010) found that the CCC-2 was better at identifying pragmatic language impairments than the Test of Pragmatic Language (TOPL-2, Phelps-Terasaki and Phelps-Gunn, 2007) in sixteen children with ASD (with an average age of 9 years) and sixteen typically-developing children (aged 6 to 10 years). In contrast, Hoffman et al (2013) found that the TOPL-2 was better at identifying pragmatic language impairments than the CCC-2 in children with William’s Syndrome aged 6 to 16 years. The TOPL-2 test has not been considered in this review of assessments as it does not assess inferential comprehension or idioms but other aspects of pragmatic communication.

Ryder et al (2008) point out the difficulties in constructing pragmatic assessments that give reliable results. They highlight that pragmatic checklists such as the CCC-2 focus mainly on the child's observed communicative behaviours and expressive language. The comprehension difficulties underlying these communicative behaviours are not assessed.

Table 1.4 below summarises the advantages and disadvantages for British children of each of the above idiom comprehension assessments.

Table 1.11 The advantages and disadvantages of existing idiom comprehension assessments

Test Name	**Advantages**	**Disadvantages**
Non-literal comprehension sub-test of the ACE	1.British 2.Standardised 3.Large standardisation sample	1.Only a few idioms are tested
Test of Language Competence, sub-test 4: idiom comprehension	1.Standardised 2.High reliability	1.American idioms 2.Level 1definition task is harder than level 2 multiple choice task
The Fullerton test for Adolescents, sub-test 8: Idiom definition	1.Standardised	1.American idioms 2.Too easy for age group assessed (11-18 years) 3.Out of print
Test of Word Knowledge	1.Standardised	1.American idioms
Idiom Comprehension Test		1.American idioms 2.Very small sample
Korean Autism Social Language Test (KASLAT)		1.Korean idioms 2.Very small sample
Informal assessment of social language and communication skills for children in primary school	1.British	1..Not standardised
Mount Wilga	1.British version 2.Standardised	1.Adult assessment 2.Very small standardisation sample (100 people aged 15-25,

		10 per year group) 3.Some outdated content 4.No cohesive theoretical basis
Understanding Ambiguity: An assessment of pragmatic meaning	1.British	1.Very small sample (15 per age group)
The Children's Communication Checklist (2nd Ed) (CC2)	1.British	1.Not standardised

The above examples demonstrate the challenge of developing an assessment of inferential comprehension.

CHAPTER ONE SECTION EIGHT: 1.8 Chapter summary and the study aims and hypotheses

1.8.1 Chapter summary

This chapter has covered: definitions of the terminology used in this study; the clinical rationale for developing an assessment of inferential and idiom comprehension; inferential comprehension in typically developing children and children with communication impairments; idiom comprehension in typically developing children and children with communication impairments; social inferencing in typically-developing children and children with communication impairments; creating a language test; current formal and informal assessments of inferential and idiom comprehension; and a chapter summary including aims and hypotheses.

The clinical justification for the need for a new British-normed assessment of inferential and idiom comprehension has been given, as there is currently a gap in this area of comprehension assessment. Existing formal and informal assessments test some of these areas but they all have disadvantages and no formal assessment currently includes sections on empathic and motivation-based inferences, as recommended by Adams et al (2009). The new test devised in this study will attempt to fill the gaps identified from the extensive literature review.

1.8.2 Study aims and hypotheses

The study aims and hypotheses, which were shaped by the current researcher's clinical practice and the review of the literature above, are:

Primary study aims and related hypothesis:

1) To develop a robust assessment of inferential and idiom comprehension for 5:00-9:11 year-old children and to provide supporting validity and reliability data on the developed assessment.

2) To provide normative and statistically significant data for inferential and idiom comprehension in typically developing children aged 5:00 to 9:11.

Related hypothesis

Null Hypothesis (Ho): The age group of the participant will have no effect on inferential and idiom comprehension scores.

Alternate Hypothesis (H1): The older age groups will have higher inferential and idiom comprehension scores.

3) To qualitatively analyse the developmental progression of inferential and idiom comprehension.

4) To carry out exploratory case studies using the HICIT with children with communication impairments to determine if it could be a useful assessment tool to assist in differential diagnosis.

Secondary study aims

The primary aim of this study was to obtain quantitative and qualitative normative data for the development of inferential and idiom comprehension in 5-10 year-old children.

Information on the gender of the participants and their socio-economic status was also obtained as part of the study. This allowed for examination of any links between gender and test performance and between socio-economic status and test performance.

Brief findings on the links between gender and language and socio-economic status and language

Lovas (2011) reports that gender difference in language development, with girls being more advanced than boys, is a frequently reported finding in the literature. Girls start to talk earlier and develop bigger vocabularies than boys in the early years. Bornstein et al (2004) reported that in the age range of one to six years, girls were more advanced in 'language performance' than boys. They suggested that this difference was due to the interaction of biological, psychological and social factors. This gender difference is small and has usually disappeared by five years of age. However, it reappears during adolescence. Boys are significantly more likely than girls to develop language impairments. Lovas (2011) reports that boys have better

verbal comprehension than girls in childhood but this advantage disappears by adolescence and she also reports that boys are better at verbal analogies than girls. These latter findings imply that boys should do better than girls on tests of inferential comprehension in the primary school years.
Spencer et al (2016) recently reviewed the findings of many studies on the links between socio-economic status and language. Law et al (2011) reported that children from areas of socio-economic disadvantage were significantly more at risk of displaying language delay than children from areas of socio-economic advantage.

Secondary study hypotheses:

1) To compare inferential and idiom comprehension between genders; to determine if there is a statistically significant difference between female and male participants.

Null Hypothesis (Ho): The gender (female or male) of the participant will have no effect on inferential and idiom comprehension scores.

Alternate Hypothesis (H1): The gender (female or male) will have an effect on inferential and idiom comprehension scores.

2) To compare inferential and idiom comprehension to the socio-economic status of the participants to see if there is a statistically significant difference between children from different socio-economic backgrounds.

Null Hypothesis (Ho): The socio-economic status of the participants will have no effect on inferential and idiom comprehension scores.

Alternate Hypothesis (H1): The socio-economic status of the participant will have an effect on inferential and idiom comprehension scores.

The next chapter, chapter two, will describe the new inferential and idiom comprehension test creation and standardisation process.

CHAPTER TWO: METHODOLOGY

The literature review in chapter 1 justified the need for a British standardised inferential and idiom comprehension assessment for primary aged children. This chapter outlines general methodological principles, and the test development and standardisation procedures. The literature review indicated that both quantitative and qualitative analysis of the test responses would provide useful information about the typical development of inferential and idiom comprehension in 5-10 year-old British children.

The new assessment will be called the Hewitt Inferential Comprehension and Idioms Test (HICIT), incorporating the current researcher’s surname.

Chapter summary

This chapter covers: the primary and secondary research aims and hypotheses; qualitative and quantitative methodologies; research epistemology; the principles of quantitative test design; and the research design of this study. The latter section (2.4) is divided into the following sub-sections: the development of the HICIT and its marking criteria; methodological procedures common to the pilot and final study; the pilot study procedures; the final study implementation; and the data analysis methods to be used in analysing the results.

Primary Study Aims and related Hypothesis:

1) To develop a robust assessment of inferential and idiom comprehension for 5:00-9:11 year-old children and to provide supporting validity and reliability data on the developed assessment.

2) To provide normative and statistically significant data for inferential and idiom comprehension in typically developing children aged 5:00 to 9:11.

   Related Hypothesis

   Null Hypothesis (Ho): The age group of the participant will have no effect on inferential and idiom comprehension scores.

   Alternate Hypothesis (H1): The older age groups will have higher inferential and idiom comprehension scores.

3) To qualitatively analyse the developmental progression of inferential and idiom comprehension.

4) To carry out exploratory case studies using the HICIT with children with communication impairments to determine if it could be a useful assessment tool to assist in differential diagnosis.

Secondary Study Hypotheses

1) To compare inferential and idiom comprehension between genders; to determine if there is a statistically significant difference between female and male participants.

   Null Hypothesis (Ho): The gender (female or male) of the participant will have no effect on inferential and idiom comprehension scores.

   Alternate Hypothesis (H1): The gender (female or male) will have an effect on inferential and idiom comprehension scores.

2) To compare inferential and idiom comprehension to the socio-economic status of the participants to see if there is a statistically significant difference between children from different socio-economic backgrounds.

   Null Hypothesis (Ho): The socio-economic status of the participants will have no effect on inferential and idiom comprehension scores.

   Alternate Hypothesis (H1): The socio-economic status of the participant will have an effect on inferential and idiom comprehension scores.

2.1 Qualitative and quantitative methodologies

Qualitative research is exploratory research, used to gain an in depth understanding of underlying reasons, opinions, and motivations for certain behaviours. It requires only a small sample of participants. It is often used in areas where theories are limited or non-existent and helps to develop hypotheses for quantitative research (Pring, 2005).

Quantitative research is based on specific hypotheses. It generates numerical data that can be transformed into interpretable statistics. It is used to quantify attitudes, opinions and behaviours (including language) and to detect patterns in the research data. Participant sample sizes need to be large so that the results can be generalised to a wider population. Structured data collection methods such as systematic observations, surveys, questionnaires and tests are used.

The development of a test for inferential and idiom comprehension therefore requires a predominantly quantitative design. However, some qualitative analysis is carried out on the responses to the test questions.

2.2 Research epistemology

In developing the study design Crotty's (1998) four main questions were addressed: The epistemology that informs the theoretical perspective; the theoretical perspective behind the methodology; the methodology that governs the choice of methods; and the methods proposed.

1. Epistemology is the theory of knowledge that defines what kind of knowledge is possible and legitimate (Crotty, 1998). It is a way of understanding and explaining how we know what we know. The epistemology of this study is objectivist, ie the approach is free and independent from particular feelings and opinions (Stokes, 2011). Objectivism purports that a meaningful reality exists independently of the mind and that we can discover the objective truth by measuring it in some way.
2. Theoretical perspective is the philosophical stance that embeds the methodology (Crotty, 1998). The theoretical perspective or 'worldview' in this study is post-positive empirical observation and measurement. This is sometimes called the 'scientific method' (Cresswell, 2014).

3. The methodology is the strategy that links the choice of particular methods to the desired outcomes (Crotty, 1998, in Feast and Melles, 2010). The methodology in this study is experimental research (Crotty, 1998).
4. The methods are the techniques used to gather and analyse data related to the research hypotheses. The methods used in this study are: sampling; measurement and scaling; statistical analysis; data reduction and content analysis (Crotty, 1998).

Pring (2005) emphasises that, in the research world, theories are temporary. They help us to develop knowledge but are rejected if they are disproved and alternative hypotheses are then proposed and tested. Hypotheses are predictions about what the researcher expects the results to demonstrate. The theories to be tested can come from anywhere (eg previous research or clinical observations). The hypotheses generated from these theories then need to be proved or disproved.

As this study is the development of an assessment of language behaviours it fits into a positivist, deductive research philosophy.

This research theory has been generated by clinical observation and the study employs a predominantly quantitative research design.

<u>Quantitative research</u>

Cresswell (2014, figure 3.4, p59) outlines the typical stages in a quantitative research study as being:

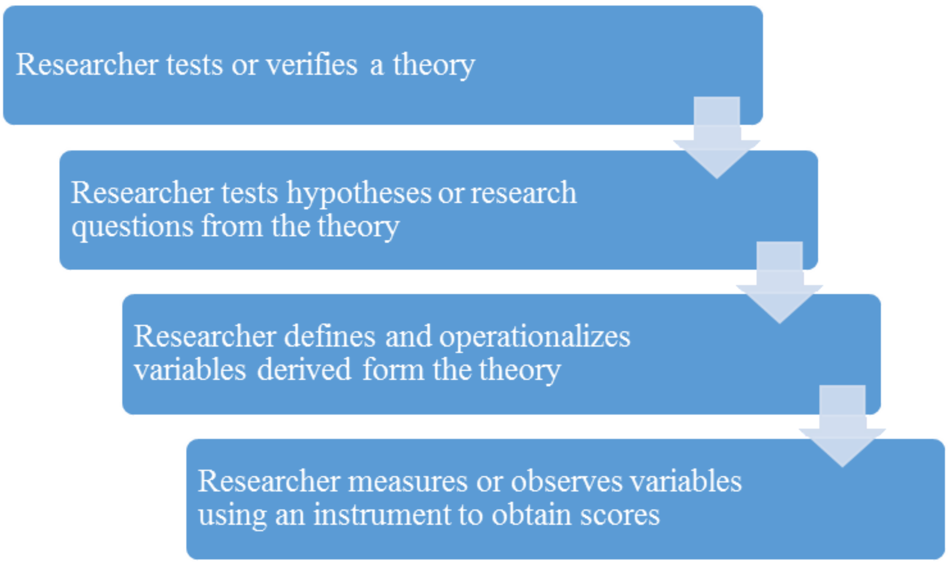


Figure 2.1 Creswell's (2014) stages in quantitative research

Creswell (2014) states that quantitative research tests objective theories by examining the relationship among variables. These variables can be measured and the numerical data produced can then be analysed using statistical tests. The researcher's theories are tested deductively. The testing should be unbiased and alternative explanations for the findings controlled for. The researcher should also report on whether or not the test findings are generalisable and replicable.

In quantitative research variables are used to form theories. Cresswell (2014) defines a variable as a characteristic that can be measured and that varies among the individuals being studied. Variables are distinguished by their occurrence in time and by their measurement. It is now accepted that cause-effect relationships cannot be proven absolutely when dealing with human studies in natural settings, so when one variable precedes another in time and it affects another variable it can only be deduced that the first variable probably causes the other. Variables often measured in studies include age, gender, socio-economic status, and specific behaviours or responses. Confounding variables are not observable during the study but may be apparent once the study is completed (eg discriminatory attitudes).

<u>Quantitative theories (Cresswell, 2014)</u>

In quantitative research variables are manipulated to answer research questions or to make research hypotheses. Theories can be stated in research proposals as a series of hypotheses, if-then logic questions or visual models. Hypotheses are used to state the direction the study will take. Research hypotheses have been selected as the way to illustrate the theories in this study. Two main forms of hypotheses are used. The null hypothesis predicts that there will be no difference between the groups being investigated. The alternative, directional, or substantive hypothesis is the researcher's prediction about the expected outcome based on prior research and, in the case of this study, also clinical observations. In some studies where the outcome cannot be specified, a third type of hypothesis, a non-directional hypothesis, is used. This study adopts the null and alternative hypotheses approach as the direction of the outcome can be predicted.

Once the study data have been collated statistical procedures need to be applied so that inferences can be drawn about the population from the study sample. The specific tests used and rationale for these is detailed in the data analysis section below. The statistical computer programme selected for use in this study was IBM's

Statistical Package for the Social Sciences (SPSS), version 19 (IBM, 2013). This computer software is now widely used in social and health sciences for the analysis of research data.

Once hypotheses have been tested and re-tested in different settings and with different populations a theory emerges which can be given a name and this can then advance knowledge in that particular field.

2.3 Principles of quantitative test design

Key issues to be addressed specifically in quantitative test design research which relate to data analysis are standardisation, sample size, validity, reliability and generalisability (Stokes and Wall, 2014, Bishop 1997).

2.3.1 Standardisation

Standardisation of a test makes it possible to compare one participant's performance with that of another participant of a similar age. For standardisation to occur the sample selected has to represent the population that the test will be used with, taking into account factors such as socio-economic differences. The sample also has to be large enough to allow an accurate measure of the average score (mean) and range of scores (variance) for each age group

Measures of central tendency include the mean (average score, ie total of scores divided by the number of participants); the median (middle score in the ranked scores) and the mode (the most frequently occurring score). Pallant (2013) recommends using the 5% trimmed mean (obtained by removing the top and bottom 5 per cent of the cases and then re-calculating the mean) if there are some extreme outliers in the data. For these outliers not to have a big effect, the mean and the trimmed mean values need to be very close.

Measures of variability illustrate the dispersion of scores and they include the range, variance and standard deviation. A frequency distribution shows how the test scores are spread out for certain norm groups. Histograms illustrate this information graphically (Gregory, 2004). The range of scores is the simplest measure of dispersion but it can be misleading when there are extreme scores ('outliers'). A way round outliers is to use the interquartile range (the 'midspread'). To calculate this, the scores are put in ascending order. The middle score is the median. The scores one quarter of the way up and then three quarters of the way up show the interquartile range. This range measurement would be applied if the median was being used as the measure of central tendency.

The variance and the standard deviation are the most used measures of central dispersion. The variance is calculated by adding up the squared differences between each score and the mean and dividing it by the number of participants (N) and the square root of the variance is the standard deviation. Standard deviation is the preferred measure in psychological assessment. Test results are usually expressed as a standard score, ie in terms of SD from the normative sample mean (Bishop, 1997). In a normal distribution most scores fall within two standard deviations above and below the mean. Fewer than 3% of the population score below 2 SDs below the mean (McCauley, 2001). Standard scores are important for clinical decision making. Many researchers consider children with scores on language tests of more than 1.5 to 2 standard deviations below the mean to require intervention.

The distribution of test scores for a large population is often depicted as a symmetrical bell-shaped curve. Skewed distributions are caused by extreme scores and they affect the mean more than the median. Positively skewed distributions, where the scores pile up at the low end of the graph, indicate that the test questions are too hard for that norm group. Negatively skewed distributions, where the scores pile up at the high end of the graph, indicate that the questions are too easy for that norm group. Normality of the assessed population can be assessed to some extent by obtaining skewness and kurtosis values. Skewness values provide an indication of the symmetry of the distribution. Kurtosis provides information about the 'peakedness' of the distribution (Pallant, 2013). Kurtosis values above zero indicate a distribution that is too peaked with short, thick tails. Kurtosis values below zero indicate a distribution that is too flat, also with too many cases in the tails (Tabachnick and Fidell, 2007). If the distribution is perfectly normal, you would obtain a skewness and kurtosis value of 0. However, this is very uncommon in social sciences research. With reasonably large samples (two hundred or more), skewness will not make a substantive difference in the analysis (Pallant, 2013). Kurtosis can result in an underestimate of the variance, but the risk is also reduced with large samples. Positively or negatively skewed scores do not necessarily indicate a problem with the scale as they often reflect the nature of the underlying construct being tested (Pallant, 2013). Many scores are not normally distributed. A significant number of scales and measures used in the social sciences have scores that are skewed, either positively or negatively.

2.3.2 Sample size

McCauley (2001) states that sampling error increases as sample size decreases. Bishop (1997) states that where normative samples are less than 50 the scores will be imprecise and results will need to be interpreted with caution.
A statistical power calculation can be used to measure how likely the study is to produce statistically significant results, depending on the size of the participant sample (Bowling, 2009).

2.3.3 Validity and reliability

McCauley (2001) reports that these two measures dominate discussions about the quality of measurement. She describes validity as being the central concept, with reliability, a less important but still vital concept, contributing towards validity.

2.3.3.1 Validity

Validity is concerned with ensuring that the instruments and methods used in the research provide accurate and dependable results, ie that the test measures what it is supposed to measure. McCauley (2001) explains that both internal and external validity are not simple concepts but are on a continuum. The most that can be said about a given measure is that it has a certain level of validity to answer a specific question regarding a specific individual's characteristics. Measures designed to assess one area would not have validity for other related areas (for example a British test of receptive vocabulary would not be valid for a non-English speaker and it would not be a valid test for comprehension of English grammatical structures). A test is valid if the inferences drawn from it are appropriate, meaningful and useful (Gregory, 2004). In psychological testing this is done by checking if the findings correlate with another similar 'gold standard' assessment (if one exists) or with findings obtained from different sources (eg for language tests for children this could be from parental or teacher reports) (Bishop, 1997). A test has face validity if it looks valid to the examinees and the examiners (Gregory, 2004). Some of the existing assessments summarised in the introduction have only face validity.

McCauley (2001) defines the three main categories of test validation as the overarching concept of construct validity with the underlying concepts of content validity and criterion-related validity (depending on the type of test). Kline (2000) emphasizes that content validity only applies to a small range of tests where the area of items is very clear-cut, such as ability and attainment assessments.

Construct validation is the accumulation of evidence to demonstrate that the test is an effective measure of the construct. When developing an assessment of a particular area of language the test developer needs to consider how the construct to be measured relates to other behavioural constructs in the world (for example age, gender, general ability, socio-economic factors, etc). The best way of measuring this construct also needs to be researched (eg pointing to pictures, describing pictures, giving verbal responses to questions etc). The two most common types of studies used to provide evidence of construct validity with children are developmental studies and group differentiation studies. The developmental method of construct validation is founded on the assumption that learning and language skills increase with age. Developmental studies can be longitudinal (measuring a single group of children over time) or cross-sectional (measuring different age groups of children at the same time). Cross-sectional studies are the most popular with test developers as the data needed for construct validation are the same as those needed to develop the norms (McCauley, 2001).

Group differentiation or contrasting group studies tests the hypothesis that two or more groups of children will differ significantly in their test results. For example, for this current test, the hypothesis would be that children with receptive language disorders, high functioning ASD or PLI would score significantly lower than typically developing children.

A third category of construct validity which is less frequently used in Speech and Language Therapy is factor analysis. It is often used in the early stages of assessment development to help narrow down the number and types of test questions. Factor analysis encompasses a range of techniques that analyse the interrelationships of a set of variables and then narrows these down to a smaller number of factors. When the target measure shares (or 'loads on') the same factor as has been demonstrated to be valid elsewhere, then construct validity is demonstrated (McCauley, 2001).

Validity tests: Exploratory and Confirmatory Factor Analysis

Factor analysis is not designed to determine if one group is significantly different from another but is a 'data reduction' technique. It looks for 'clumps' among the inter-correlations of a set of variables (Pallant, 2013). The analysis is frequently used by test developers to refine their large number of related variables to a more manageable number.

Hatcher (2013) defines Exploratory Factor Analysis (EFA) as a multivariate procedure used to determine the number and nature of common underlying (or latent) factors (or variables) that underpin the observed variables. These common factors are hypothetical constructs.

Confirmatory Factor Analysis (CFA) comes under the Structural Equation Modelling (SEM) umbrella, a flexible set of procedures allowing researchers to test theory-driven models that assume there are causal relationships between the variables (Hatcher, 2013). CFA produces indices that demonstrate if the model constructed is a 'good fit' or not. It also produces factor loadings and inter-factor correlations. It is used later in the research to confirm specific hypotheses concerning the structure underlying a set of variables (Pallant, 2013).

Content validity checks that the content of the test is consistent with the constructs it is aiming to measure. This encompasses content coverage and content relevance. Some of this is covered in the process of developing the test. For norm-referenced assessments decisions need to be made about the complexity of the constructs to be tested as well as the numbers and kinds of items to be used. The HICIT is a norm-referenced assessment constructed of eleven sub-tests assessing different areas of inferential comprehension and one sub-test assessing comprehension of idioms. McCauley (2001) states that as many as one and a half to three times as many test questions are written for the original version as will be expected to be used in the final version. The items are then tried out in a pilot study and following this the performance of each item is assessed using item analysis (which can include factor analysis). Any unsatisfactory items are then discarded or modified, leaving the most valid items in the final test. The difficulty of each item (p) is then calculated by dividing the number of children who answered it correctly by the number of people who answered the question. A p of 1.0 would mean that the test questions was answered correctly by every child and a p of 0.0 means that the test questions was not answered correctly by any child. A *p* value of 0.5 is the ideal value, but a range of 0.3 to 0.7 is considered acceptable. Items that fall outside of this range should be omitted or modified.

Item discrimination conveys the extent to which children perform similarly on the item as they do on the total test. The most popular ways of measuring item discrimination are with an item discrimination index or an item-total test score point biserial correlation (McCauley, 2001).

2.3.1.2 Reliability

Reliability is a major factor affecting validity. A measure can only be valid if it is also reliable. The reliability of a scale indicates how free it is from random error (Pallant, 2013). The larger the reliability of a test, the smaller the error. The more items in a test, the more likely it is to be reliable.

There are three main types of reliability: a) internal consistency reliability, b) interrater reliability and c) test-retest reliability (McCauley, 2001).

a) Internal consistency reliability looks at how dependable the assessment results are. It examines the consistency of the test content or how well the test items hang together (Pallant, 2013). It measures how much different items in the test correlate with the overall score, or the degree to which the items that make up the scale are all measuring the same underlying attribute. The same data that provide the test norms are used to compute the internal consistency. The most common method of calculating this was originally by generating a split-half reliability coefficient where the test performances of a group of participants were compared for two halves of the measure. Usually even items are compared with odd items. This method has since been made more sophisticated and tests of homogeneity such as the Kuder-Richardson formula (KR20, in Hatcher, 2013) or the coefficient alpha are now more frequently used, particularly in tests for children.

The most commonly used statistical measure of internal consistency reliability is Cronbach's coefficient alpha. This provides an indication of the average correlation among all the items that make up the scale.

Values range from 0 to 1, with higher values indicating greater reliability. Nunnally (1978, in Pallant 2013) recommends a minimum level of 0.7 Cronbach alpha values, dependent on the number of items in the scale. Pallant (2013) says values above 0.7

are acceptable, but values above 0.8 are preferable. Garson (2013) reports that, although the widely-accepted social science Cronbach alpha cut-off for a set of items to be a scale is usually 0.7 or higher, some criteria are as lenient as 0.6.
McCauley (2001) states that a value of less than 0.2 is a slight, negligible correlation; 0.20-0.40 is a low correlation; 0.4-0.7 is a moderate correlation; 0.7-0.9 is a high correlation; and over 0.9 is a very high correlation. It is important to stress, however, that a strong correlation between factors is not evidence of a causal relationship between them (McCauley, 2001). However, for scales of less than ten items Cronbach alpha values can be quite small (Briggs and Cheek, 1986, in Pallant, 2013).
Reliability can also be considered in terms the degree of error affecting an individual score. The statistical calculation of standard error of measurement (SEM) is used for this. It is calculated by multiplying the standard deviation of the test by the square root of one minus the reliability coefficient. Reliability increases as the SEM decreases. The SEM is used to obtain a confidence interval around specific scores (McCauley, 2001).
Reliability can be a problem when assessing children, especially those with attentional difficulties, as for children the shorter the test the better (McCauley, 2001). Few behavioural measurements are one hundred percent reliable as many subject, test and situational factors affect test results. Measurement errors can arise during item selection, test administration and test scoring. Children's performance in language tests can be influenced by many factors such as tiredness, concentration lapses, emotional state and guessing responses.
A reliable assessment would minimise these factors as much as possible. Classical test theory (CTT) assumes that: measurement errors are random, the average error of measurement is zero, true scores and error scores are uncorrelated, and errors on different tests are not correlated (Gregory, 2004). It is assumed that an observed score is the sum of the true score plus some non-systematic error (McCauley, 2001).
Item response theory (IRT), in particular Rasch analysis, is now replacing CTT as the preferred model for assessment development. A recent study by Jabrayilov et al (2016) proved that IRT was superior to CTT in detecting individual change as long as the test contained at least 20 items. However CTT was better at detecting change in individuals in shorter tests.

Reliability measure. Item Response Theory (IRT) - Rasch Analysis

Rasch analysis is a type of IRT test. Eckes (2015) finds the Rasch measurement approach invaluable for analysing language assessment responses. The dichotomous Rasch model is used for analysing binary data (ie scores of 0 or 1). It measures the degree to which the test scoring is justified by the data collected, ie that the total score means the same across the compared groups. A mathematical calculation computes the probability that an examinee answers a particular item correctly depending on the difference between the ability of the examinee and the difficulty of the item (Eckes, 2015). When the examinee's ability (latent trait) is equal to the difficulty of the item there is a 0.5 probability of a correct response. The probability of a person responding correctly to a question with lower difficulty than their ability is greater than 0.5 while the probability of them responding correctly to a question with greater difficulty than their ability is less than 0.5.

If the invariance of responses across different groups of people does not hold, then taking the total score to characterise a person is not justified. However, data do not always fit the model perfectly. If this happens, it is important to find where the misfit is worst and try to understand it in terms of the construction of the items and the understanding of the variable in terms of its theoretical development.

There are some problems with the Rasch model, however. These are summarised by (Kline, 2000) as:

a. It assumes that items are equally discriminating, but it is extremely difficult to construct any test in this way.
b. It assumes that the participants do not guess some of the answers which is extremely unlikely.

b) Inter-rater reliability is the extent to which two or more assessors agree on scoring for the same population (Bowling, 2009). It is also called interexaminer-, interscorer-, interobserver- and interjudge-reliability. It is a measure of consistency across testers. It compares the test results carried out by a range of examiners on the same children in the same conditions. Ideally the testers score the tests live or from a tape or video. It is computed by two or more examiners scoring up the assessment on the same participants independently. Their test scores are then correlated.

The formula for this is:

$$\text{Percentage Agreement} = \frac{\text{Number of agreed scores}}{\text{Number of judgements}} \times 100$$

This can be calculated in different ways, including using a Kappa test. An inter-rater reliability score of less than 0.40 indicates poor agreement; 0.40-0.59 fair agreement; 0.60-0.74 good agreement; and 0.75-1.00 excellent agreement.

c) Test-retest reliability or test stability checks if the same results would be replicated on the same participants again after 2-4 weeks. This test-retest interval is particularly important in children's testing because natural developmental changes will occur over longer periods (McCauley, 2001). However, the problem with re-testing in such a short time frame is that there may be some carry-over effects from the initial assessment. This could inflate or deflate the second test results (due to practice effects or remembering and repeating incorrect responses). Kline (2000), on the other hand, recommends that the testings should be separated by at least three months to be trustworthy and the samples should comprise at least 100 participants. The reliability is calculated by a test-retest reliability coefficient between the two sets of scores. The higher the reliability coefficient the greater the degree of relatedness between the repeated scores. Kline (2000) states that a test-retest correlation of 0.8 is the minimum figure required.

Murphy and Davidshofer (2005) highlight the difficulties with establishing the reliability of tests. They state that, in an ideal world, the more reliable the test, the better. However, in the real world, the most reliable test might be too long, too expensive to purchase, or too difficult to administer. If tests are being used as a preliminary screen the test reliability is less important than if it is being used to make critical decisions about the examinee.

McCauley (2001) stresses that the validity and reliability of tests need to be interpreted in the context of who is being tested and for what purpose. Kline (2000) states that, as proving validity is such a multi-faceted and complex procedure, relatively few tests have good evidence for their validity.

2.3.1.3 Generalisability

This refers to the degree to which the conclusions from the research apply to other settings and similar contexts. Positivist, deductive style studies, such as the current one, use large sample sizes and quantitative and statistical analyses so that the results can be applied to a wider population than just the sample (Stokes and Wall, 2014).

2.4 The methodology for this study

The order of implementation of the different components of this study was: the development of the pilot test, converting the pilot test to the final test, development of the test marking criteria, the devising of procedures common to both the pilot and final study, carrying out the pilot study, and conducting the final study.

2.4.1 Test development

The development of the Hewitt Inferential Comprehension and Idioms Test (HICIT) was the first task to be completed in the study, as participants could not be recruited until there was a test to be trialled. Key assessment areas were drawn from the review of existing assessments of verbal reasoning, the literature review and resources used for intervention in these areas. The questions used for the HICIT were based on specific areas identified in the above but the researcher generated her own specific questions for each section. They were not copied verbatim from existing assessments. Many of the simpler inferential comprehension questions were based on the Blank (1978a) language of learning questions. They also fit in with Filiatrault-Veilleux et al (2015) story grammar questions (eg 'What is the problem?'; 'How does X feel?'; 'What does X want?'; 'What could X do?'; 'What do you think will happen next?').

The original study proposal was to develop an inferential comprehension assessment for Key Stage 2 primary children aged 7 to 11 years, as documented in the MMU ethics forms (appendices iv and v). However, the literature review indicated that some inferential comprehension abilities were already developing in reception age children (aged 4-5 years) so the lower boundary of the proposed test was moved down to 4:06 years. The literature review also revealed that some understanding of idioms had developed by 5-6 years of age, so idiom comprehension was added to the test.

The key areas initially selected were:

2.4.1.1 Inferential comprehension

- Making simple deductions
- Answering why questions (justifying predictions, justifying a decision, justifying an idea, explaining the means to a goal, explaining the construction of objects, explaining the logic of compound words, explaining obstacles to an action) (Blank et al, 1978a)
- Predicting events (what will happen if ….?)
- Formulating solutions to problems (what could you do….?)
- Identifying the causes of an event
- Making inferences from an observation (how do you know….?)
- Multiple choice inferences from spoken scenarios; making inferences from a verbal scenario
- Making inferences from short stories.

According to Cain and Oakhill's (1999) and Botting and Adams' (2005) definitions of sub-types of idioms, the inferences included in the short stories sub-section of the HICIT are predominantly gap-filling/ bridging inferences, where new information is related to old or the children need to integrate their own general knowledge with information in the text to fill in details that are not explicitly stated. The inferences made are also 'offline', ie made retrospectively by the child (Saladana and Frith, 2007).

2.4.1.2 Social inferencing

- Taking someone else's perspective
- Interpreting body language and facial expression (how does X feel?);
- Theory of Mind emotions questions (Situation-based; desire-based and belief-based), based on (Howlin et al, 1998) and (Hadwin et al, 2015). The current researcher used Howlin et al's (1998) categories but provided her own examples of the emotions scenarios.
- Mental state verbs (Dennis et al, 2001, Spanoudis et al, 2007)

- Strange Stories based on (Happe, 1994, Kaland et al, 2005, O'Hare et al, 2009). The current researcher generated her own strange stories which fitted in to some of Happe's (1994) story categories.
- Faux pas based on (Baron-Cohen et al, 1999) and (Zalla et al, 2009). The final faux pas sub-section contained 3 amended faux pas based on Baron-Cohen et al's (1999) scenarios and one novel story (the lady in the clothes shop). The current researcher used faux pas examples from her own experience which were similar in type to some of Baron-Cohen et al's (1999) faux pas.

2.4.1.3 Idioms

These were compiled from a range of assessments and resources. For example: Don't Take it so Literally (Legler, 1991); Black Sheep Press, Pragmatics 12 (idioms) and Pragmatics 3 (speech bubbles) (www.blacksheeppress.co.uk); and Super Duper Publications, Idioms Fun Deck. (superduperinc.com).

One hundred idioms were narrowed down by the researcher to fifty of the most common UK idioms. These were trialled on two British teenagers and a British adult aged 13, 16 and 48 years. Any idioms rated as 'difficult' or 'uncommon' by these participants were removed. If two idioms were considered similar in meaning only one was selected. This left 38 common British idioms.

Norbury (2004) conducted a pilot study on 100 typically–developing children aged 11-14 to obtain the idioms to use in her study of idiom comprehension in communication-impaired children. The children were read out 52 idioms in a neutral sentence context (eg Mum told me to 'pull my socks up'), 26 of which were transparent and 26 opaque. The children were asked to define each idiom and rate how familiar it was on a scale of 1 to 5. Definitions were scored on a 3 point scale ranging from 0 (don't know or literal answer) to 2 (correct). Norbury then deliberately chose 10 low familiarity idioms to use in her study.

Caillies and Le Sourn-Bissaoui (2008) developed normative data for their idiom study by asking 200 French adults to use 6 point scales to rate 653 idiomatic expressions on the dimensions of familiarity (from 'never heard before' to 'very well-known'), compositionality/transparency ('individual words are not related' to 'are closely related to their figurative meaning') and literality (from 'implausible' to 'very plausible'). They selected French idioms that had both a literal and an idiomatic meaning. They then chose 20 idioms that varied in compositionality (10 transparent and 10 opaque) but that had high levels of familiarity and literality for their study.

Cain et al (2009) point out that the choice of tasks used to measure idiom comprehension may influence the results. Idiom explanation tasks may disadvantage younger children due to their less well developed expressive language and comprehension skills compared to older children. The preferred method of assessing idiom comprehension is, therefore, the use of a multiple-choice task to test idioms in context.

Kerbel and Grunwell (1998, in Norbury, 2004) and Norbury (2004) both used definition tasks when testing idiom comprehension in children with communication impairments. Norbury (2004) acknowledged that this type of response format can be difficult for both typically-developing and communication-impaired children. However, she justified this choice as multiple-choice tasks which include literal foils are too easy or too confusing for children. She suggested that a definition task would give a better insight into the child's thought processes when trying to define idioms. The current researcher therefore uses an idiom definition task in the HICIT. She also deliberately selected the most familiar British idioms as the test was aimed at younger children than those in the above studies. She does not separate idioms into transparent and opaque and presents the idioms in isolation, not in context. Also she uses the scoring system of acceptable (1) or unacceptable answer (0) with a detailed list of acceptable responses provided. This is easier to implement than Norbury's (2004) scoring system.

2.4.1.4 Interpreting pictorial information (visual-verbal inferencing)

The selected resources were 'What's Wrong Colorcards' and 'Fun Pictures' (speechmark.net). These are visual absurdities that have to be explained verbally.

2.4.1.5 Miscellaneous components

The selected components included verbal absurdities, detecting the irrelevant information in passages and sentence completion inferences.

The response mode for this assessment was carefully considered. Setting Yes/No questions requires excessive amount of testing given the 50/50 chance element (Botting and Adams, 2005). Offering a choice of picture responses also introduces a chance element (1 out of 4 if there are 4 pictures to choose from). Filiatrault-Veilleux (2015) report that visual support for verbal questions is most effective for assessing inferential comprehension in pre-school children (3-6 year-olds). However, the majority of the children in this study are over 6 years of age. It is also very difficult to represent all desired inference questions pictorially. In addition it is problematic to convey accurate information about nonverbal communication (facial expression, gestures etc) in static pictures and the same picture could be interpreted in different ways. In their study, Bishop and Adams (1992) found that some incorrect answers on their inferential comprehension test were due to the children misinterpreting the picture information. For example, in story B, some children thought the policewoman was a nurse and thought the story was about a hospital visit. They concluded that if the child misinterprets the pictures, it affects their understanding of the whole story. This study also demonstrated that children with language impairments had the same level of inferential comprehension difficulties with and without picture supports to the stories. Pike et al (2010) reported that illustrations both facilitated and interfered with inferential comprehension on a written bridging-inferences task.

Consequently, it was then decided to develop a purely verbally-based rather than a picture based assessment, ie the child would be asked to listen to and answer verbal questions from the tester. This, therefore excluded section 2.4.1.4 above.

However, supporting pictures and the written word were used for sections 7, 8 and 11 to support the child's auditory memory. Stylised facial expression pictures were used for 7 and 8 ('happy', 'sad', 'angry' and 'scared' for section 7 and just happy and sad for section 8 see Appendix vii). The written words: Yes/No/Maybe were presented before section 11, mental state verbs (Appendix viii).
The selected mode of child response was also verbal, ie expressive language. Some researchers, eg Kerbel and Grunwell (1998), claim that definition tasks underestimate children's comprehension. However, the justification for using

definition answers in this test is that picture assessments can overestimate comprehension level by giving the child a 25% chance or 33.3% chance of getting the correct answer if they have to select 1 out of 4 or 1 out of 3 pictures respectively (Cain et al, 2009). Also, the level of expressive language needed to score correctly on the test questions is basic. Most correct answers can be single words or a short phrase or sentence. For example:

> **Section 2 Why Question 10:** 'Why shouldn't you agree to take a lift in a car from people you don't know?', a sufficient answer would be 'Might take you away'.
>
> **Section 4 Formulating solutions inferences Question 11**: 'How could you find out what your teacher looked like when she was a little girl?' The answer 'A photo' is sufficient.

Adams et al (2001) state that a minimum standard of expressive ability is also required for the inferential comprehension sub-test of the ACE assessment. Finally, the populations of children with high level comprehension impairments targeted by this assessment are those with social communication, pragmatic or specific receptive language difficulties. These children usually have good functional expressive language skills, more than adequate to answer the questions posed in this assessment.

Cultural factors were taken into account in the assessment. As for the ACE test (Adams et al, 2001), only questions using basic world knowledge were included and any items which might present problems for specific cultural groups were not included.

The sections were honed down further from the original sections i) to iv) above to make the first version of Pilot Study Test, (Pilot HICIT 1). A summary of all the versions of the test is given in chart form at the end of this section (Table 2.1).

This pre-pilot version of the HICIT was trialled on 5 children known to the author, aged 5 to 13 years. It took 40-55 minutes to administer. Clinical experience suggests that this is too long for a standardised assessment as most initial assessments sessions in clinics are only allocated one hour and this test would only assess

receptive and not expressive language. To reduce the length of the test any similar questions were omitted, along with any questions the 13 year-old had difficulty explaining.
The final Pilot Study Test (appendix ix) consisted of a total of 210 questions. See table 2.1 at the end of section 2.4.2 for a summary of the sections included in this pilot test.

Wording of questions was carefully selected. In section 4, 'Formulating Solutions', the question 'What *could* you do if ….?' is used rather than 'What *can* you do if…? as the former gives more possible answers. It is also preferable to 'What *should* you do if….?', as this could imply to the child that a moral decision rather than a practical solution is required.
In section 5, 'Explaining Inferences', due to the frequent semantic confusion between ask and tell in young children, it was decided to use the question form 'How do you know..?' rather than 'How can you tell that….?'. Also, the term 'know' links more clearly into the child's theory of mind skills.

2.4.2 Converting the pilot assessment to the final HICIT

The results section (chapter 3) documents which questions were omitted from the final assessment and why, based on the statistical analyses carried out on the pilot study data.
The Final HICIT Assessment was reduced to 12 sections from 14 and the number of questions in each section was reduced where possible to lessen further the test administration time. The number of questions was narrowed down from 210 to 140, a reduction of a third. The 'Why Questions' section was collapsed into one big section rather than having the a) to d) sub-sections within it.
Within each section the questions were put in order from easier to more difficult to understand. This was based on the overall number of correctly answered questions per section across the whole pilot group Consequently the Final Assessment questions are in a very different order to the Pilot Assessment.
Table 2.1 below summarises the process of test question reduction.

Table 2.1 A summary of the test question selection process

Section number and name	**First iteration (initial trial version)**	**Second iteration (pilot version)**	**Third iteration (standardisation version)**
	Number of questions in section	**Number of questions in section**	**Number of questions in section**
1)Simple Deductions	10	10	8
2) Why questions:	(Total of 24)	(Total of 20)	12
a) Explaining obstacles to action and experience	6	5	
b) Cause-effect/ Prevention	6	5	
c) Justifying a decision	6	5	
d) Construction of objects	6	5	
3) Making Predictions	11	8	6 (sections 3 and 7 from pilot combined)
4) Formulating Solutions	24	21	12
5) Explaining Inferences	30	22	12
6) Making Inferences from Short Passages	16	16	10
7) Taking Other Perspective	10	8	
8) Situation-Based Emotions	16	16	16 (final section 7)
9) Desire-Based Emotions	8	8	
10) Belief-Based Emotions	24 (12 x 2)	8	8 (final section 8)
11) Mental State Verbs	14	14	12 (final section 9)
12) Strange Stories	20	16	12 (final section 10)
13) Faux Pas	24	18	12 (final section 11)
14) Idioms	38	25	20 (final section 12)
Total	**269**	**210**	**140**

See appendix x for a copy of the final HICIT form. The sections are given in table 2.1 above.

The test is not intended to be a vocabulary test so lower frequency words such as 'enamel' and 'pedestrian crossing' can be explained if the child asks. In section 2 the 'Why Questions' 11 and 12 are presented in the following way on the test response form:

11. Why are teeth covered in enamel not cotton wool?

(NB If the child asks what enamel is you can say: 'The stuff on the outside of teeth')

12. Why do we need pedestrian crossings on roads?

(NB If the child asks what a pedestrian crossing is you can say: 'Like a zebra crossing').

The test is also not intended to tax auditory memory so as many repetitions as were needed of any question could be requested. Interestingly however, Paris and Upton (1976) and Oakhill (1984) (both in Adams et al, 2001) found that developmental changes in the ability to infer were not due solely to increased memory capacity.

The pilot study was carried out purely by the researcher so the administration instructions are minimal. The final test instructions are detailed and specific to allow other testers (including MMU SLT students who carried out some of the final assessments) to carry it out in a standardised manner. These are given below:

> *Instructions to testers*
> *Start by telling the child: 'I have got lots of questions to ask you. You may not know all the answers and it is fine to say 'I don't know' if you do not know the answer to any question'.*
> *Say that some questions have more than one right answer.*
> *Tell the older children (year 3 onwards) that a lot of the questions might be very easy for them.*
> *Tell the child you can repeat each question as many times as he/she needs you to.*
> *If the child gives an idiosyncratic response (eg Section 2, Q6 'Why can't you read in the dark?': 'Because I didn't eat my carrots', Q7 'Why are windows made out of glass and not out of bricks?': 'The Georgians had bricks in their windows to avoid paying window tax'; Section 4, Q11 'How could you find out what your teacher looked like when she/he was a little girl?: 'Go back in time'; Section 5, Q3 'How do you know that someone is angry?': 'In a cartoon they would have steam coming out of their ears') prompt them with 'Yes, but what about in real life/ nowadays?'*
> *If you think the child has partially answered a question, prompt with 'Anything else?' NOT with 'Why?*
> *For example: Section 2, Q4 Tester 'Why mustn't you play with matches?': Child 'Because they're dangerous'; Tester 'Anything else?'*
>
> *Section 5 'How do you know that....?' If the child answers 'Because they tell you' for any of these questions prompt with 'Yes, and how else do you know?'*
> *Question 8. 'How do you know that someone has got toothache?' If the child replies 'they go to the dentist' repeat the question stressing the 'you' (Yes, but how do YOU know they have got toothache?)*
> *This is a verbal test so keep the test form out of sight as much as possible so the child does not try to read the questions.*
> *Write down the child's exact verbal response to each of the following questions. If he/she uses gesture to answer some questions describe the gestures used.*

2.4.3 Development of the test marking criteria

Letts and Leinonen (2001) point out that there can be a wide range of acceptable answers to inferential comprehension questions responses and that children can go through a logical inference process and come up with an answer that is different from the expected ones.

Adams et al (2009) carried out a pilot study of their Assessment of Comprehension and Expression 6-11 sub-test questions on twenty children with typically developing language and 20 children with language impairments. The aim for the Inferential Comprehension sub-test was to decide on the range of acceptable and unacceptable responses, as decided by a panel of five experienced speech and language therapists. Answers were considered to be unacceptable if they were unrelated to the question, related but imprecise or they demonstrated a lack of inferential understanding. A final exhaustive list of acceptable and a list of unacceptable responses was constructed which was then developed into a scoring guide used for the final assessment. This is given in a table in their test manual, p37-38. For their items 1-4 the scores are 0 for an unacceptable response and 1 for an acceptable response. Items 5-8 elicit more elaborate responses so are scored 0 for an unacceptable response, 1 for an acceptable response and 2 for two or more acceptable responses from their table. Item 9 gets a score of 2 if the child produces an acceptable response which includes justification and 0 if not. Filiatrault-Veilleux et al (2015) support Adams et al's (2009) 0-2 point classification of the types of responses given by children, claiming it leads to a more nuanced discrimination of a child's performance. However, the current researcher considered that 1 acceptable answer should be sufficient to get maximum marks. She therefore opted to stick with the simpler 1/0 (correct/ incorrect) scoring system using very comprehensive and wide-ranging marking criteria for what count as acceptable responses.

The total number of test questions in the final HICIT is 140 so the maximum total score possible is 140.

The acceptable answers for each section of the test vary from single word, one correct response only, to short phrases or sentences with many acceptable answers per question. This is represented in table 2.2 below:

Table 2.2 Summary of the types of answers required in the HICIT

Single word answers. Only 1 correct response. Choice of 2 (happy/sad)	Single word answers. Only 1 correct response: Choice of 3: (Yes/No/ Maybe)	Single word answers. Only 1 correct response. Choice of 4 (happy, sad, angry, scared)	Single word answers. Multiple acceptable responses	Short phrase/ sentence response. Multiple acceptable responses
8. Belief emotions	9. Mental state verbs	7. Situation emotions	1. Deductions 11. Faux pas	2. Why Questions 3. Predictions 4. Solutions 5. Inferences 6. Passages 10. Strange Stories 12. Idioms

The range of acceptable scores expanded over the course of the Pilot and Final study as children introduced new examples. Local dialectal and colloquial expressions were fully accepted. For example ‘narky’ for ‘annoyed’, ‘they’re shiny on their head’ for ‘sweating’, ‘get a wiggle on’ for ‘get your skates on’. Even some unique creative terms were allowed, eg Section 2, Q10: ‘Why shouldn’t you agree to take a lift in a car from people you don’t know?’ A 5 year-old child’s response ‘cos they might stranger-danger you.’ See appendix xi for the final marking criteria.

.

Markers are asked to use their clinical discretion when marking any test question answers that are not covered by the detailed marking criteria. The introduction to the marking criteria reads:

Please use your clinical discretion to mark the few answers that may not be covered by the marking criteria below. Mark any justified responses as correct.
For example:
Section 4, question 7: ‘What could you do if you burnt a cake you had baked for your mum’s birthday?’
Child: ‘Give it to her because my mum likes burnt cake’
Also – accept local dialectal vocabulary. For example, ‘narky’ for ‘moody’, ‘get a wiggle on’ for ‘get your skates on.’

2.4.3.1 Examples of the development of the marking criteria

Single word answers

1. Simple Deductions (8 questions)

Questions 2, 3 and 6 have only 1 acceptable answer. Questions 1, 4, 5, 7 and 8 have a few possible answers (eg number 7 could be violin, viola or fiddle).

7. Situation-based emotions (16 questions)

The instructions given at the beginning of this section are outlined below:

Use the 4 emotions symbols for this section. Before you start check that the child knows what picture depicts what emotion by asking him/her to point to each one as you say them.
Instructions to the child: 'For the next section, choose the feeling that you think fits best from: happy; sad; angry; or scared. There might be two possible answers for some of these but choose the one feeling that fits the best.'

It is acknowledged that there is a 25% probability of the child selecting the correct answer by chance.

8. Belief-based emotions (8 questions)

The instructions given at the beginning of this section are outlined below:

Just have the 2 symbols for happy and sad visible for this section (ie fold the emotions sheet in half).
Instructions to the child: "For the next section, choose the feeling that you think fits best from: happy or sad"

The researcher acknowledges that there is a 50% chance of the child getting this answer correct by chance. This is addressed in the results and discussion section.

9. Mental State Verbs (12 questions)

The instructions given at the beginning of this section are outlined below:

Use the page with the words Yes, Maybe and No on it. Point to the appropriate word as you give the instructions below.
'I am going to tell you about 12 different people and a door. If you think the door is definitely locked say 'yes'. If you think the door might or might not be locked say 'maybe'. If you think the door is definitely NOT locked say 'no'.

The researcher acknowledges that there is a 33.3% chance of the child giving the correct response by chance.

11. Faux Pas (12 questions)

As this is not a test of auditory memory the instructions below are included at the beginning of this section:

N.B It is OK to remind a child who specific people are if they ask, eg: Q1- If the child asks 'Who was Karen again?' Tester: She's the one who said: 'Did you see how Jill was dancing?'

This section has 4 scenarios which have 3 questions each: How does X feel?; How does Y feel: and 'Did X know that Y……?' These are presented in alternating order across the sections.

For example: Disco: It was the school disco. Jill was in one of the toilets. Karen and Sara came in to the toilets afterwards. Karen said: 'Did you see how Jill was dancing. Wasn't she terrible!' Then Jill came out of the toilet.

1) How does Karen feel?

2) How does Jill feel?

3) Did Karen know that Jill was in one of the toilets?

The hardest questions to answer is 1), the one involving second order theory of mind. It is also acknowledged that 'embarrassed' is a low frequency and later acquired emotion adjective. However, more high frequency adjectives were also accepted (see below).

The initial correct responses allowed were:1) embarrassed 2) angry/upset 3) No. However, due to the children's varied responses the following acceptable answers were added:

1. embarrassed/guilty/ashamed/full of shame/awful/terrible/regretful/wanting the ground to swallow them up/bad/mean/rude/sorry/wanting to take back what they said/ worried that Jill might have heard.
2. upset/angry/mad/annoyed/sad/disappointed/offended/shocked/grumpy.

<u>**Questions requiring more than one word answer**</u>

2. Why Questions (12 questions).

Unfamiliar vocabulary can be explained if the child asks (eg 'enamel' in question 11 and 'pedestrian' in question 12).

3. Making Predictions (6 questions)

4. Formulating solutions (12 questions)

Q4: What could you do if you wanted to eat your soup but it was too hot?

The initial correct responses were 'blow on it' or 'wait for it to cool down'. However other creative solutions from the children were added in, such as add milk/cream/cold water, dip some bread in it, fan it, put it by the window, put it in the fridge.

Q10: How could you talk to somebody who lives hundreds of miles away?

The initial right answer to this question was just ‘on the phone’, but new technology has necessitated a much wider range of correct responses including electronic communication such as via webcam, Skype, Face Time, Facebook, e-mail, Twitter etc). In addition, as e-mail and twitter were included then other forms of written communication had to be too (eg a letter).

5. Explaining Inferences (12 questions)

This section had to be prefaced with the instructions to further probe all ‘Because they tell you’ answers to ensure an opportunity was given for the child to give one of the identified correct responses.

Q.3 How do you know that someone is angry?

The initial correct responses to this question were ‘because they shout/frown/go red in the face’. Again, the children suggested many more suitable alternatives such as pull a cross face, have lines on their forehead, their eyebrows go together, clench their fists, gnash their teeth, stamp their feet, they take it out on someone else).

Q.5 How do you know that someone is too hot?

The initial right answers were: they sweat, go red in the face. The children added the following responses: they pant, fan themselves, put a fan on, open the window, take some clothes out, they’re shiny on their head.

6. Making inferences from short passages (10 questions)

Q.3 Ellie paid the money to get in. She got changed and put her clothes in a locker. She put the key band on her wrist and took her towel with her.

What was Ellie doing?

The original answer to this was ‘going swimming’ but ‘going to the gym’ was added later.

Q.4 Alisha and her mum went to see a concert but had to come back home without seeing it.

What happened?

This question had an almost endless list of possible correct answers:

They couldn’t get a ticket; they lost their tickets; the concert was cancelled/called off; it was sold out; the concert had already finished/they missed it; there was a fire at the concert; they didn't take enough/any money for the tickets; they were too late to get in; they went on the wrong day; 1 of them (or the band members) were sick/ hurt/injured; they got stuck in traffic/delayed/their car broke down; Alisha was too young to get in; there were too many people there; bad weather prevented it (eg a storm).

Q.6 A group of friends went out to a party together. They had planned to get the bus home afterwards but ended up taking a taxi.

Why?

The modern day answer 'because it was cheaper to get a taxi for that many people' had to be added in to the list of possible responses.

Q10. Dad bought a chicken to cook for Sunday dinner but the family ended up eating just vegetables and gravy.

What happened?

The unexpected answer 'they had decided to become vegetarian' was added to the list of acceptable responses.

10. Strange Stories (12 questions)

Each section apart from the 'misunderstanding' had 2 questions attached to it. The first was a Yes/No question about the truth of the person's statement. This means there is 50% chance of getting the correct answer by chance. The second was a 'Why did X say that?' question.

The Misunderstanding Story had 4 questions attached to it:

Jack hurt another child on purpose at school, but no teachers saw him do it. At the end of the day he started to walk home. A teacher ran after him shouting 'Stop, Jack!' Jack turned round and said 'I'm really sorry for doing it Miss'. The teacher only wanted to give him his lunch box that he had forgotten.

5) Why did Jack say 'I'm really sorry for doing it Miss'

6) Did the teacher want to tell Jack off?

7) How did the teacher feel when Jack said 'I'm really sorry for doing it, Miss' (If the child gives the wrong answer to the above say: 'I think she was surprised.' Then ask

8) Why?

Contrary Emotions story: James' group of friends is going swimming on Saturday morning. James really wants to go with them but he knows that Sam, a big boy who bullies him at school goes swimming in the same pool on Saturday mornings. When his friends ask him if he wants to go swimming with them he says 'No.'

9) Is it true what James says?

Both a 'Yes' and 'No' response were accepted for this question as it emerged over the course of the trial that both were acceptable answers. This means that the effect of Q9 is negated in the final score.

12. Idioms (20 questions)

The following introduction is given to this section:

Instructions to the tester: Say all of these idioms with a fairly neutral tone of voice and facial expression and do not gesture, so that you do not give away the meaning. Introduce the section by giving the following example:

"Some sayings mean something different from their words. For example 'give me a hand' doesn't mean 'chop a hand off and give it to me'. It means 'help me please.' What do the following sayings mean? It is okay to say 'don't know' if you do not know what it means."

This is a similar script to Norbury's (2004) from her study of idiom comprehension in language-impaired children, but the HICT wording was developed independently by the current researcher:

"I'm going to ask you some funny expressions people sometimes use. You know when people say 'it's raining cats and dogs' they don't mean cats or dogs are in the sky, they just mean it's raining really hard. Don't worry if you don't know these expressions, just think about the words and what they could mean."

The children were encouraged to have a guess if they did not know what the idiom meant. Some of the acceptable answers are other idioms. For example: 'he's having a ball' for 'he's having a whale of a time'; 'winding me up' for 'driving me round the bend'; 'got my hands full' for 'all tied up at the moment'.

Some answers will be locally acceptable answers (eg 'get a wiggle on' for 'get your skates on'). It is not possible for the researcher to be aware of all of these so the examiners must use their discretion when scoring such responses and this is written in to the examiner's instructions.

2.4.4 Methodological procedures used in both the pilot and final study

Key issues to be addressed in this project are sampling (including gender, socio-economic and cultural considerations), exclusion criteria and ethical considerations. Examples of the specific forms used in the study are given in the pilot or final study procedural sections.

Sampling

The most rigorous method of selecting a norm population is through stratified random sampling. This is where the population is classified on variables such as age, sex, educational level, social class and race and then an appropriate percentage of participants is randomly chosen from each section (Gregory, 2004).

Ideally, for the findings of the research to apply to the population studied, the participants should be drawn randomly from the population. Without random sampling the external validity of the research is reduced and the findings are not generalizable to that population. However, practical difficulties mean that pure random sampling is not always possible (Bowling, 2009).

The sampling in this study was convenience or purposive sampling. That is, the participants were recruited from a known population that was easily accessible (Bowling 2009). The participants for the pilot study and the final study were recruited from primary schools near to where the researcher or the student researchers lived.

Socio-economic considerations

Many studies have linked lower socio-economic status to lower performance on standardised language tests (Spencer et al, 2016). Therefore, for the final study, it was important to gain a sample of participants from a wide range of socio-economic areas.

Schools were asked to provide information about the percentage of children who received free school meals in the school. The Free School Meal Entitlement (FSME) is a measure of low income but also provides a crude measure of socio-economic status (Hobbs and Vignoles, 2007). It is a proxy rather than a direct measure and is used frequently in educational research and policy. It is acknowledged that the FSME is not fully robust as it measures income only, does not recognise the changing eligibility for or past history of receiving FSM and does not take into account other aspects of deprivation (NIAR, 2010). However it is considered the best proxy measure when compared to alternatives such as geographical deprivation indices (NIAR, 2010).

The national average of free school meal entitlement (FSME) in English schools in 2012 was 19.3% in maintained nurseries and state funded primary schools and 16% in state funded secondary provision (DfE, 2012).

Exclusion criteria

As the study aimed to gain norms for British language comprehension, the children required for the study were typically-developing children whose first language was English. Children with significant special needs (eg children with a statement of Special Educational Needs), those receiving Speech and Language Therapy (SLT) input and those whose first language was not English were excluded from both the pilot and final study. A statement of Special Educational Needs is an English formal educational document that details a child's learning difficulties and the help that he or she will be given. It was replaced by the Education, Health and Care Plan (EHCP) in September 2014. Norbury (2004) included only native British English speakers who had no disorders of language and cognition and were not receiving any special educational support in her idiom pilot study to collect idiom norms. Huber-Okrainec et al (2005) excluded children who had a diagnosis of language disorder, a requirement for special educational services or who were educated in a language other than English, in their study of idiom comprehension. Cain et al (2009) excluded children who had a statement of special educational needs and whose first language was not English from their idiom study on 40 children in the North West of England.

Bishop (1997) warns that there is a danger of missing important information about cause and effect when applying exclusion criteria. However, she supports the careful use of exclusion criteria in research to ensure that an appropriate sample of the selected population is being tested. Children who were getting some extra support with literacy or numeracy on 'school action' or 'school action plus' (which are the steps below a statement of educational need) were included in the current study to prevent a population bias in favour of more able children only.

2.4.4.1 Ethical considerations

All studies should be carried out within an appropriate ethical framework to protect the participants and the researchers from potential harm and to make the research findings more robust (Stokes and Wall, 2014). Accepted research practices and codes of conduct must be followed. The British Psychological Society (BPS) Code of Human Research Ethics (BPS, 2014) describes some of the main ethical considerations in research projects as obtaining informed consent from the participants, ensuring and safeguarding confidentiality of data and the anonymity of the participants, and respecting privacy. The BPS also outlines the four main general ethical principles to be adhered to as: being respectful, competent, responsible and behaving with integrity. The standards of conduct and professional behaviour from the professional and regulatory bodies were also adhered to. These are the Royal College of Speech and Language Therapists (RCSLT) and the Health and Care Professionals Council (HCPC).

The Manchester Metropolitan University (MMU) ethical guidelines were followed to ensure that all ethical principles were adhered to and the project received ethical approval from the university (see appendices iv and v).

Obtaining informed consent: Gaining informed consent is essential before any study can commence. This ensures that the participants' rights will be protected during the data collection (Cresswell, 2014). In this research, schools' Head Teachers or Special Needs Co-ordinators (SENCOs) were used as gatekeepers to allow access to the children in the schools. Gaining their permission to carry out the study in the school was the first step in the process. This was achieved by sending the Head Teacher an introductory letter explaining the study aims.

As both the pilot and final study involved child participants, parental consent was obtained (Neil, 2005). The school sent out information letters about the project to parents which included an opt-out consent form. Swanell (2014) points out that the human research ethics guidelines allow opt-out consent if the research is low risk (ie the only foreseeable risk is one of discomfort), of significant public benefit or involves a dataset so large that individual consent is impractical to obtain. The current research fits into the 'low risk' category. Parents were informed that even if they gave consent for their child to take part they could withdraw him or her from the study at any time.

Verbal assent was gained from the child prior to each assessment in school. The children were also informed at the beginning of the assessment (or before each section if it was felt appropriate for that child) that it was fine to answer ‘don’t know’ if they were unsure of an answer and that they could discontinue the assessment and go back to class at any time. This was to minimise any anxiety they may have had about the assessment.

Safeguarding confidentiality of data: Completed scoring sheets contained only the child’s initials and year and month of birth. Each child was allocated a study number. This pseudo-anonymization was necessary as parents were given the option to withdraw their child’s data at any time from the study. Each child therefore had to be traceable. Completed forms were kept in a secure, locked location. In the final study written permission was gained from a sub-set of randomly selected parents to audio record their child’s assessment session for purposes of moderation (appendix vi). These recordings were kept securely on a password protected computer.

Respecting privacy: This was more difficult to ensure in the school settings as space was very limited and the children sometimes needed to be assessed at a table in a school corridor or cloakroom. However, the children were not disclosing any personal or confidential information, so ensuring privacy during the assessments was not essential.

2.4.5 Carrying out the pilot study

Participant recruitment: It was originally planned to recruit one-hundred-and-forty primary aged children aged 4:06 to 10:06; 10 girls and 10 boys from each age group (4:06, 5:06. 6:06. 7:06, 8:06, 9:06, 10:06). However a pre-pilot test carried out on one child from each age group indicated that the majority of the assessment was too difficult for the 4:06 year-olds and too easy for the 10:06 year-olds. It was decided to trial a few of the easier sections of the assessment with four 4:06 year-olds and a few of the harder sections with eight 10:06 year-olds to establish floor and ceiling levels. In addition, the pilot assessment took 50 minutes to administer for each child. The original target of 140 children was therefore revised down to 62 children due to time constraints. These children were recruited from 2 primary schools (1 junior and 1 infant) in the researcher's locality in the North West of England. The final numbers assessed were:

Age groups 5:06, 6:06, 7:06, 8:06 and 9:06 – 10 children in each group, 5 girls and 5 boys (50 in total)

Age group 4:06 – 2 girls and 2 boys (the boys were identical twins and the girls were non-identical twins) (4 in total).

Age group 10:06 – 4 girls and 4 boys (8 in total).

A letter was sent out to four local primary school heads outlining the study (appendix xii) and asking them to contact the researcher if they were interested in participating. As no response was received from these letters they were followed up with phone calls to the Head Teachers. Two schools approached for the pilot study declined to participate due to impending OFSTED inspections. One junior school and one primary school in the North West of England were finally recruited. The Head Teachers identified the Special Needs Co-ordinator (SENCO) as the contact person for the researcher in each school.

Free School meals (FSM) Index: The junior school that participated in the study had a FSM percentage of 9 and the infant school a FSM percentage of 0. Both schools therefore had a below average number of children on FSM, indicating a higher socio-economic status than average for these schools participating in the pilot study.

The school sent out information letters about the project to parents, which included an opt-out consent form (appendix xiii).

Procedure: A few weeks after the forms went out to parents the researcher met with the school Special Needs Co-ordinator (SENCO) to exclude any children with significant special educational needs, those receiving speech and language therapy and those whose first language was not English. In addition, those children whose parents had sent back the form saying they did not want their child to participate in the study were excluded. Certain classes were identified as the targets for the study and the teachers were informed about the study by the researcher. Specific days were selected for the researcher to come into each school. The pilot study took 45-50 minutes to administer. The younger children needed this to be split into two separate sessions as their attention spans lasted 25-30 minutes. It was only possible to see 6-7 children per day due to break times, lunch times and specific teaching sessions that the children were not allowed to miss.

Prior to the assessment each child was asked for their assent using the following wording:

> *"Hi, my name is ----- and I am a Speech and Language Therapist. Do you know what that job is? It is to help children who find it hard to understand and talk. I know you don't have any difficulties with talking. This test has been made for children who have problems with talking, but first it has to be tried out with children who don't have any problems, like you. Are you happy to help me with this?"*

Only 1 child out of 263 in the whole study (pilot and final) did not give assent so was not assessed.

The child was asked each question and his/her answers were recorded verbatim on the assessment form.

To allow more time for administering the assessments in the school day the completed tests were scored up at the end of the day using the marking criteria devised by the researcher.

Equipment: The equipment used in the pilot study was the pilot assessment forms, a pen, the marking criteria and stickers to give out as rewards for those children who wanted one.

Data Analysis Methods: The SPSS programme (version 19, 2013) was used to analyse the data.

Descriptive statistics: The total test mean scores by age and gender; mean scores by age for each section; and the range of scores for each section for each age group were calculated.

Spearman's correlation is used to explore the strength of the relationship between two continuous variables. A positive correlation indicates that as one variable increases, so does the other. A negative correlation indicates that as one variable increases, the other decreases (Pallant, 2013).

Spearman's correlation test was applied to the data. The results were then used to decide which items to omit.

Reliability: Cronbach's coefficient alpha was applied to assess internal reliability before and after the removal of the items.

2.4.6 Carrying out the final study

2.4.6.1 Participant recruitment: A further 200 children were recruited, mostly from schools in the North West but also from other areas of the country. In their pilot study for the Assessment of Comprehension and Expression 6-11 (Adams et al, 2009) the 20 typically developing children tested did not then go on to participate in the main data collection. The 50 children in the current researcher's pilot study were included in the final study (using their initial data converted into the final HICIT format), making a total of 250 children tested.

The researcher carried out 122 final HICIT assessments in two primary schools in the North West and eight Manchester Metropolitan University (MMU) student volunteers in their final year of study carried out 78 final HICIT assessments, some in the North West and some in other areas of England. More details about the numbers, age groups and locations in which they tested are given in the following results chapter (chapter 3). Cain et al (2009) used 25 undergraduate students from Lancaster University to collect data for their idiom study in the Northwest of England.

The sample size in this study is smaller than that of some other published language tests. For example the ACE standardisation was carried out on 117-145 children in each age group (6:00-11:11), compared to 50 children per age group in the current study. However, taking into account the time constraints of carrying out a part-time PhD study and the majority of the assessments being carried out by the researcher, 250 is a realistic sample size.

The relatively large sample size of 250 in this study ensured that any skewness and kurtosis effects would not significantly affect the analysis (Pallant, 2013).

2.4.6.2 Free School meals (FSM) Index: The researcher used 2 primary schools in the North West, 1 with a Free School Meals (FSM) percentage of 9.6 and one with a FSM percentage of 53.6. Students from MMU carried out assessments in schools in their hometowns. The percentages of FSMs in these schools ranged from 4.8 to 50.3.

2.4.6.3 Procedure: The researcher requested involvement from the schools in the North West using the letter in appendix xiv. The students were given the letter in appendix xv to send to their local school. The procedure was the same as for the Pilot study. The researcher briefed the staff in her schools at twilight staff meetings. The student volunteers were sent procedural instructions to follow (see appendix xvi). They were asked to send the completed forms back to the researcher for her to score. However, they were also asked to mark one of the completed assessments themselves, using the marking criteria provided. This provided a measure of student-researcher inter-rater reliability.

The children were briefed in the same way as for the pilot study. In addition, 10% of these children were audio-recorded for moderation purposes. The parents of the children selected for audio-recording had sent in a written consent form agreeing to this (appendix vi). The children were asked verbally for their assent prior to the recording. No child refused to be audio-recorded.

2.4.6.4 Equipment : The equipment used in the final study comprised: the final assessment forms; a digital tape recorder; a pen; the marking criteria; stickers to give out as rewards for those children who wanted one.

2.4.6.5 Challenges to data collection in the pilot and final study:

Space was a premium in all schools so it was often difficult to get a designated quiet room in which to do the assessments. Some of the researcher’s testing had to be carried out in noisy school corridors or large cloakrooms where there were frequent distractions and interruptions. Locating the children to assess was occasionally problematic as the classes sometimes moved to a different location for certain lessons and activities (eg PE, IT, assembly). The length of the assessment session was challenging for some of the younger children as their attention span was shorter than that of the older children. Consequently, the assessment sessions for the 4 year-olds and some of the 5 year-olds were split into two separate sessions throughout the day with the child having a rest in between the sessions. When scoring up the assessments there were two examples of a missing unrecorded response from two different children. The researcher was able to go back to the setting the day afterwards to obtain these missing data.

2.4.6.6 Data Analysis Methods

Descriptive statistics: The following were calculated for the total final scores by age and gender, and for the section scores by age: the mean, the 5% trimmed mean, the median, the variance, the standard deviation, the range, the interquartile range, the skewness, the kurtosis, and the Kolmogorov-Smirnov significance value.

Inferential statistics: Parametric tests are used where data is drawn from a normally distributed population; there is homogeneity of variance (data are drawn from populations with approximately equal variances); and the data are measured on an equal-interval scale. Non-parametric tests are distribution-free and no assumptions are made about the data. Parametric tests are more powerful than non-parametric tests (Pring, 2005). As the variables in this assessment are measured in interval scales then parametric statistical tests are used (ie T-tests, Analysis of Variance (ANOVA) (McCauley, 2001).

A one-way Analysis of Variance (ANOVA) was carried out to explore the effect of socio-economic status on the total test scores. A two-way ANOVA was used to explore the effects of gender and age group on the total test scores.

Validity: Exploratory and confirmatory factor analyses were carried out.

Reliability: Internal reliability was assessed by applying Cronbach's alpha and carrying out Rasch analysis of the data. Rasch is a type of Item Response Theory test analysis. Jabrayilov et al (2016) would recommend the use of this over Classical Test Theory (CTT) as the HICIT contains over 20 items. External reliability was tested by measuring inter-rater reliability.

2.4.6.7 Qualitative analysis of the HICIT responses

The main data analysis in this study is quantitative. However, although the responses are scored correct/incorrect (1 or 0), many questions allow a wide range of correct responses. Norbury and Bishop (2002) developed nine categories of incorrect responses for their qualitative typology of inference error types. These were: failure of literal comprehension, a wrong inference, a wrong immature inference, an immature reference, an odd inference, a 'because he did' answer, a 'scope' answer (along the right lines, but too specific or too vague to be correct), lack of expressive ability, and no response. The current researcher finds these categories too vague or overlapping. Consequently she developed the very detailed marking criteria specifying a wide range of possible correct answers. However, there will still be occasional answers that necessitate the use of the tester's own judgement. Some qualitative data on the test responses are presented at the end of the results chapter.

2.5 Chapter summary

This chapter has covered: the primary and secondary research aims and hypotheses; qualitative and quantitative methodologies; research epistemology; the principles of quantitative test design; and the research design of this study. The latter section (2.4) is divided into the following sub-sections: the development of the HICIT and its marking criteria; methodological procedures common to the pilot and final study; the pilot study procedures; the final study implementation; and the data analysis methods to be used in analysing the results.

The following results chapter will commence with a review of the study aims and hypotheses and will then give a summary of the results of the quantitative and qualitative analyses described above.

CHAPTER THREE: RESULTS

Overview of the chapter

This chapter restates the study aims and hypotheses. It then outlines the pilot study results and converting the pilot assessment to the final assessment. Subsequently tests of validity (exploratory and confirmatory factor analysis) and reliability (Cronbach's Alpha, Rasch analysis and inter-rater reliability) are reported. A summary of the final study results is then provided. This is sub-divided into: participant information, descriptive statistics and inferential statistics. The inferential statistics test the hypotheses that age, socio-economic factors and gender affect the HICIT scores. Qualitative analysis of the results is detailed in the penultimate section. Four exploratory studies using the HICIT with children with communication impairments are summarised in the final section.

Primary Study Aims and related Hypothesis:

1) To develop a robust assessment of inferential and idiom comprehension for 5:00 to 9:11 year-old children and to provide validity and reliability data supporting the developed assessment.

2) To provide normative and statistically significant data for inferential and idiom comprehension in typically developing children aged 5:00 to 9:11.
   Related Hypothesis
   Null Hypothesis (Ho): The age group of the participant will have no effect on inferential and idiom comprehension scores.

   Alternate Hypothesis (H1): The older age groups will have higher inferential and idiom comprehension scores.

3) To qualitatively analyse the developmental progression of inferential and idiom comprehension.

4) To carry out exploratory case studies using the HICIT with children with communication impairments to determine if it could be a useful assessment tool to assist in differential diagnosis.

Secondary Study Hypotheses

1)To compare inferential and idiom comprehension between genders; to determine if there is a statistically significant difference between female and male participants.

Null Hypothesis (Ho): The gender (female or male) of the participant will have no effect on inferential and idiom comprehension scores.

Alternate Hypothesis (H1): The gender (female or male) will have an effect on inferential and idiom comprehension scores.

2)To compare inferential and idiom comprehension to the socio-economic status of the participants to see if there is a statistically significant difference between children from different socio-economic backgrounds.

Null Hypotheses (Ho): The socio-economic status of the participants will have no effect on inferential and idiom comprehension scores.

Alternate Hypothesis (H1): The socio-economic status of the participant will have an effect on inferential and idiom comprehension scores.

<u>3.1 Pilot study results</u>

Sixty-two children aged 4:06 to 10:06 were assessed with the pilot version of the HICIT. There were ten children in the age groups 5:06-9:06 (5 boys and 5 girls), eight children aged 8:06 (4 boys and 4 girls) and four children aged 4:06 (2 twin boys and 2 twin girls). Table 3.1 summarises the mean overall section scores and table 3.2 shows the range of scores in each age group from4:06-10:06. An X in a box denotes that that section was not administered to that age group either because, after trialling it on one or two children, it was judged to be too difficult for some of them (the 4:06 year-olds) or too easy for others (the 10:06 year-olds). Table 3.3 summarises the mean total test scores for the 5:06-9:06 year-olds.

3.1.1 Mean overall section scores by age group for 4:06 to 10:06 year-olds

There were ten children (N=10) in the age groups 5:06 to 9:06. The 4:06 year (N=4) and the 10:06 year (N=8) age groups only completed certain sections not the whole test. X denotes the sections that were not completed with these age groups.

Table 3.1 Mean overall pilot HICIT section scores for 4:06–10:06 year-olds

Age	N	1 /10	2 /20	3 /8	4 /21	5 /22	6 /16	7 /8	8 /16	9 /8	10 /8	11 /14	12 /16	13 /18	14 /25
4:06	4	4.5	11.25	4.75	13	X	X	5.5	X	6.75	X	X	X	X	X
5:06	10	7.4	16.3	6.7	18.7	15.7	9.9	7.5	9.5	7.7	4.8	7.2	11	10.5	2.2
6:06	10	7.2	17	7.2	18.9	19.2	13.1	7.8	11.1	7.1	5.6	8.4	11.8	10.4	4.4
7:06	10	8.6	17.4	7.4	19.4	19.2	14.2	7.9	12.9	8	6	9.6	13.2	11.9	7
8:06	10	9.3	18.5	7.7	20.6	19.5	15.2	8	14.1	8	7.7	11.7	13.7	13.7	8.9
9:06	10	9.4	19	7.8	20.3	20.9	14.9	8	15.5	8	7.6	11.2	14.1	15.5	15.3
10:06	8	X	X	X	X	X	X	X	X	X	X	12.1	14.4	14	15.9

3.1.2 Range of scores for each section for 4:06 to 10:06 year-olds

Table 3.2 Range of pilot HICIT section scores for 4:06–10:06 year-olds

Age	N	1 /10	2 /20	3 /8	4 /21	5 /22	6 /16	7 /8	8 /16	9 /8	10 /8	11 /14	12 /16	13 /18	14 /25
4:06	4	4-5	6-15	6-15	3-7	X	X	11-17	X	5-8	X	X	X	X	X
5:06	10	5-9	12-19	5-8	14-21	12-18	5-13	6-8	9-11	5-8	3-7	4-9	9-15	6-17	0-7
6:06	10	5-10	13-19	5-8	10-21	14-21	7-15	7-8	9-14	4-8	3-8	5-10	7-15	7-12	1--8
7:06	10	5-10	14-19	6-8	16-21	16-22	10-16	7-8	10-15	8	4-8	6-13	10-15	11-14	1-14
8:06	10	7-10	17-20	7-8	19-21	15-22	13-16	8	11-16	8	6-8	10-14	10-16	11-16	1-16
9:06	10	9-10	16-20	7-8	18-21	18-22	12-16	8	14-16	8	7-8	9-14	11-16	13-18	12-19
10:06	8	X	X	X	X	X	X	X	X	X	X	10-14	12-16	12-17	12-20

3.1.3 Mean total test scores by year group (ages 5:06 to 9:06 only)

Table 3.3 Mean total test scores by year group (ages 5:06 to 9:06)

N = 10 per age group (5 boys and 5 girls).

Average Age	Total Mean score Boys	Total Mean score Girls	Total combined Mean score /210
5:06	129.6	140.6	135.1
6:06	157.6	140.8	149.2
7:06	157.4	168	162.7
8:06	173.4	179.8	176.6
9:06	188	187	187.5

As noted in the previous chapter, section 2.4.2, the results from this pilot study informed the design of the final version of the HICIT. The following section outlines how statistical analyses were used to help to convert the pilot HICIT to the final version.

3.2 Converting the pilot HICIT to the final HICIT

The results of the Spearman's correlation were used to decide which items in each section of the pilot test should be taken out. Any questions with a negative, or very low or zero correlation (the question was answered correctly by all age groups) were removed. Spearman's correlation is used in order to explore the strength of the relationship between 2 continuous variables. A positive correlation indicates that as one variable increases, so does the other. A negative correlation indicates that as one variable increases, the other decreases (Pallant, 2013).

Cronbach's Alpha (CA), the most commonly used statistical measure of reliability, was computed for each pilot section before and after removal of these questions. Details about CA are given in the previous methodology chapter. The results are summarised in table 3.4 below.

Table 3.4 Cronbach's Alpha scores before and after item deletion

Pilot section number. The final test section no. is given after -> symbol	**Questions deleted**	**Items before/ after**	**CA before**	**CA after**
1 Deductions	1, 5	10 / 8	.561	.597
2 Why Qs	2,3,5,7,8,12,13,15	20 / 12	.540	.573
3 Predictions	1,2,6,8	8 / 4	.207	
7 Perspective	1,2,3,6,7	8 / 2	.131	
3 & 7 amalgamated ->3		16 / 6		.575
4 Solutions	3,4,7,10,11,12,13,15,17	21 / 12	.735	.765
5 Inferences	3,4,7,9,12,13,17,19,20,21	22 / 12	.682	.697
6 Passages	3,6,9,13,14,15	16 / 10	.770	.765
8->7 Situation emotions	None	16 / 16	.732	.732
9 Desire emotions	Whole section deleted			
10->8 Belief emotions	None	8 / 8	.602	.602
11->9 Mental state verbs	10,14	14 / 12	.683	.744
12 ->10 Strange stories	5,6,7,8	16 / 12	.683	.721
13->11 Faux Pas	1,2,3,13,14,15	18 / 12	.738	.603
14 ->12 Idioms	1,4,5,18,19	25 / 20	.893	.883

3.2.1 Justification for item or section removal

Section 9 – Desire-based emotions. All of the participants were close to ceiling on this section so the whole section was removed.

All of the questions in sections 8 and 10 were kept as they were as they were all needed to maintain the pattern of questions.

Sections 6, 13 and 14 – The CA value went down slightly after the removal of some questions, however, the 'after CAs' were still over 7, 6 and 8 respectively. These sections needed to have questions removed to reduce the administration time of the final test.

The weakest CA scores were for section 2 (.573), the next was amalgamated sections 3 and 7 (.575), the next was section 1 (.597).

Borderline CA scores were for sections 10 (.602), 13 (.603) and 5 (.697).

Good CA scores were for sections 12 (.721), 8 (.732), 11 (.744). 4 and 6 (.765).

A very high CA score was found for section 14 (.883).

As the pilot sample was so small it was decided to include all of the reduced sections except for section 9 in the final assessment, even though some had a weak CA score.

3.2.2 Validity tests: exploratory and confirmatory factor analysis

3.2.2.1 Exploratory Factor Analysis (EFA)

EFA was carried out on the 200 children with a complete data set (6:00 to 9:11 year-olds). The results are presented below:

Table 3.5 Exploratory factor analysis results (6:00-9:11 year-olds). Total variance

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.564	46.368	46.368	5.564	46.368	46.368
2	.942	7.849	54.217			
3	.803	6.691	60.908			
4	.747	6.224	67.132			
5	.691	5.761	72.893			
6	.652	5.436	78.329			
7	.532	4.435	82.764			
8	.527	4.394	87.159			
9	.458	3.819	90.978			
10	.394	3.286	94.264			
11	.377	3.143	97.407			
12	.311	2.593	100.000			

Extraction Method: Principal Component Analysis.

Table 3.6 Exploratory factor results (6:00-9:11 year-olds). Correlation matrix

Correlation Matrix

		Deductions	Why Qs	Predictions	Solutions	Inferences	Passage Qs	Sit Emotion	Belief Emot	Ment State	Strange Story	Faux Pas	Idiom
Correlation	Deductions	1.00	.48	.37	.39	.40	.48	.35	.30	.31	.43	.42	.48
	Why Qs	.48	1.00	.40	.46	.49	.48	.38	.23	.30	.39	.48	.46
	Predictions	.37	.40	1.00	.43	.48	.55	.29	.24	.32	.46	.31	.43
	Solutions	.39	.46	.43	1.00	.45	.43	.39	.29	.25	.46	.36	.37
	Inferences	.40	.49	.48	.45	1.00	.60	.42	.26	.35	.57	.48	.58
	Passage Qs	.48	.48	.55	.43	.60	1.00	.44	.33	.38	.56	.50	.57
	Sit Emotion	.35	.38	.29	.39	.42	.44	1.00	.33	.38	.39	.43	.58
	Belief Emot	.30	.23	.24	.29	.26	.33	.33	1.00	.29	.30	.25	.39
	Ment State	.31	.30	.32	.25	.35	.38	.38	.29	1.00	.33	.30	.40
	Strange Sto	.43	.39	.46	.46	.57	.56	.39	.30	.33	1.00	.46	.54
	Faux Pas	.42	.48	.31	.36	.48	.50	.43	.25	.30	.46	1.00	.58
	Idioms	.48	.46	.43	.37	.58	.57	.58	.39	.40	.54	.58	1.00

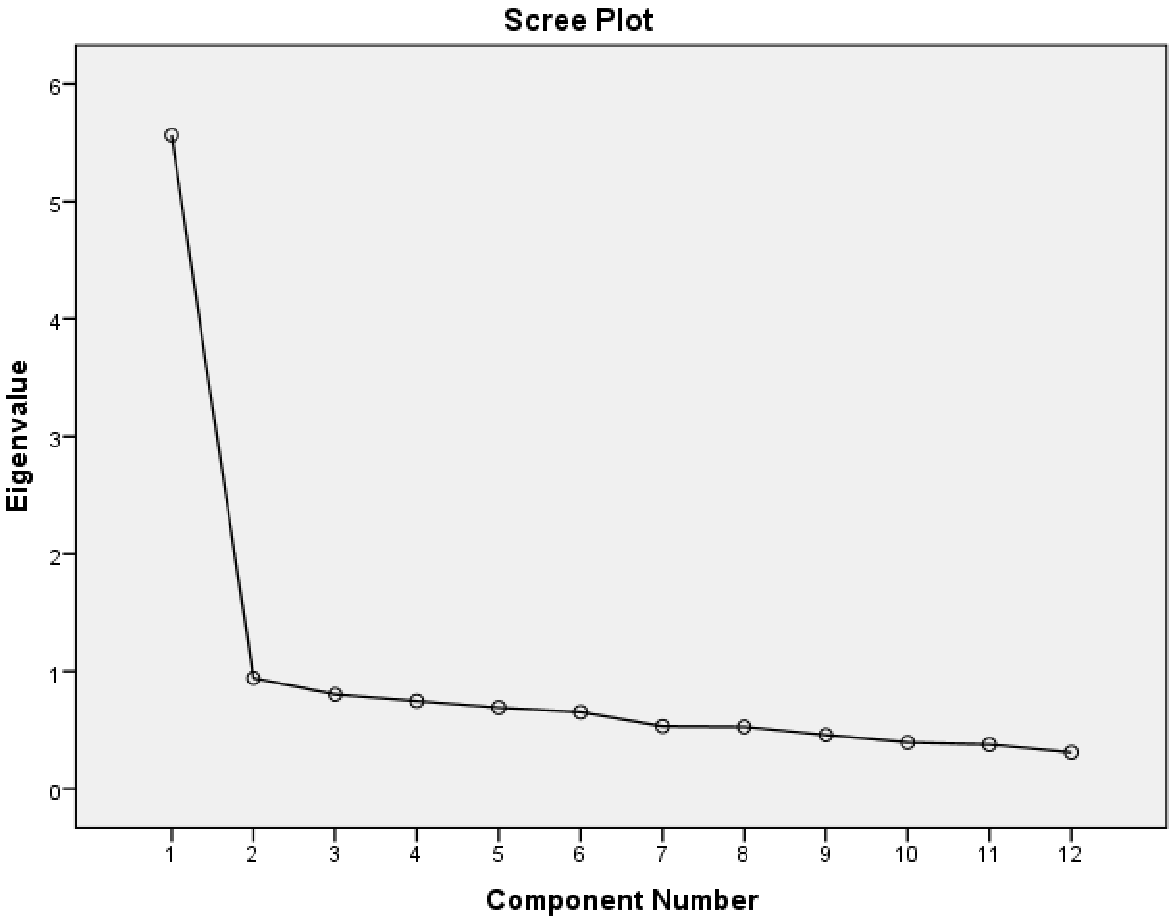


Figure 3.1 Scree plot (6:00-9:11 year-olds)

Table 3.7 Component matrix

Component Matrix[a]

	Component
	1
Idioms	.79
Passage Qs	.79
Explain Inferences	.76
Strange Story	.73
Faux Pas	.69
Why Questions	.69
Deductions	.66
Situational Emotions	.66
Predictions	.65
Solutions	.64
Mental State Verbs	.55
Belief Emotions	.49

Extraction Method: Principal Component Analysis. A 1 component extracted

Interpretation of these results

To verify that the data set is suitable for factor analysis the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) has to be over 0.6, the significance value of the Bartlett's test of Sphericity has to be 0.5 or less and the correlation matrix has to show at least some correlations of $r = .3$ or greater. The values came out as .93 and .000 respectively, and the majority of the correlations in the matrix are more than .3 so the data set is suitable for factor analysis (Pallant, 2013). The correlations under .3 in the matrix in table 3.6 above are highlighted in yellow. Most of these occur in the Belief Based Emotions sub-section which has a 50/50 chance of being answered correctly from a choice of two answers (happy or sad).

To determine how many factors there are, any components having an initial Eigenvalue of 1 or more on the Total Variance Explained table are extracted. Only the first component has an Eigenvalue of above 1 (5.56, highlighted in blue in table 3.5 above). This is also demonstrated on the scree plot (figure3.1) where only the first factor is above 1. This indicates that there is one main factor underlying all of the inferential comprehension and idiom sub-tests.

The Component Matrix table 3.7 above shows the unrotated loadings of each of the items on the 1 component. A value of above .4 is considered a quite strong loading (Pallant, 2013). All of the items load above .4 so the 1 factor solution is appropriate. The weakest loadings are on the mental state verbs and the belief based emotions sub-tests (.55 and .49 respectively).

3.2.2.2 Confirmatory Factor Analysis (CFA)

CFA was carried out on the data from the 200 children with complete data sets (6:00 to 9:11 year-olds). The results are presented below:

Standardised estimates

Standardised factor loadings range from -1.00 through 0.00 to +1.00, with values closer to +1.00 representing a stronger effect. The indicators with the largest factor loadings are considered to be the best measures of the underlying factor (Hatcher, 2013).

The figures below demonstrate the CFA results with all 12 sub-tests included, with sub-section 8 omitted and with sub-section 8 omitted and e7 and e12 co-varied.

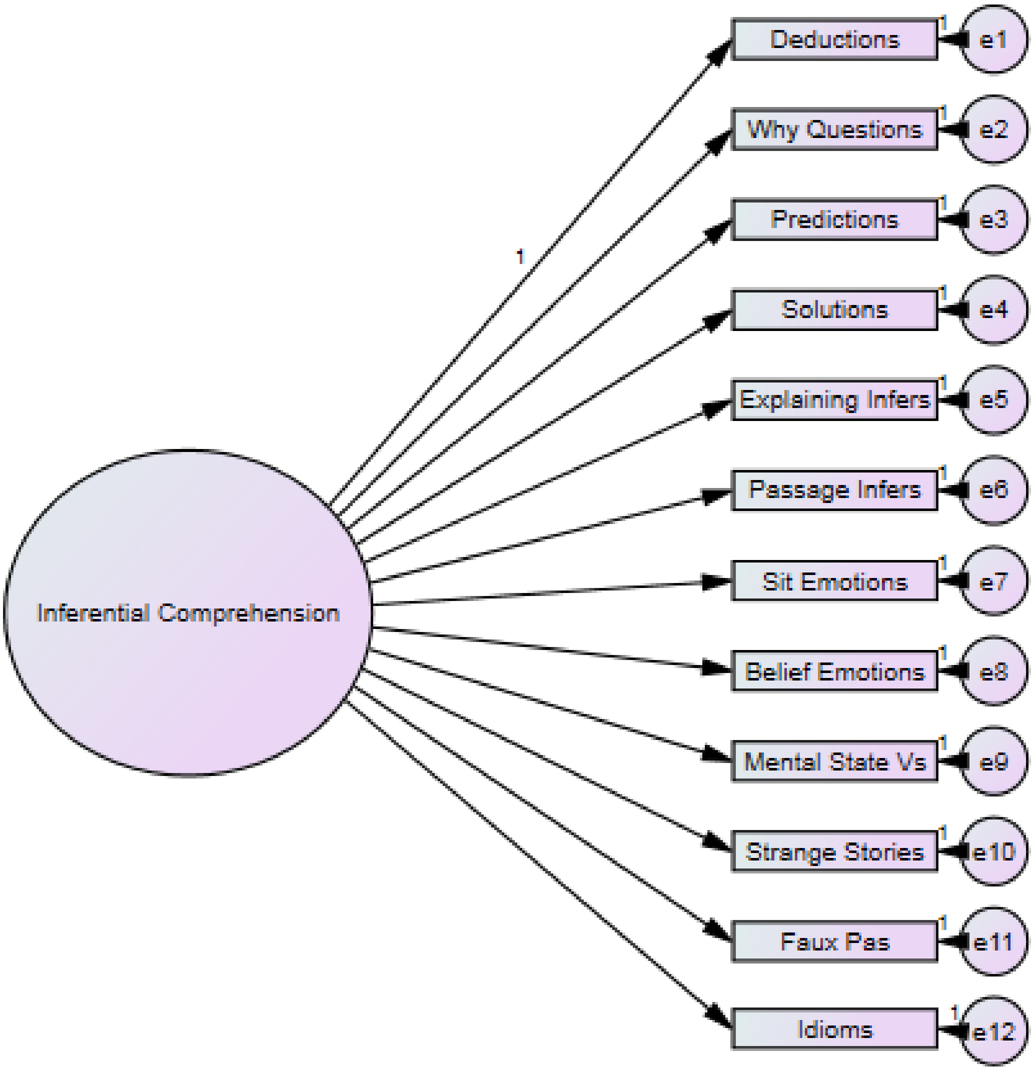


Figure 3.2 Confirmatory factor analysis with all 12 sub-tests included

In this CFA the loadings range from .44 (belief-based emotions) to .77 (idioms and passages). The table below demonstrates the loadings in descending order from strongest to weakest: The final column shows the results from the Exploratory Factor Analysis (EFA) component matrix in green.

Table 3.8 CFA standardised estimates and comparative EFA component matrix values for all 12 sub-tests

Sub-test	**CFA Standardised Estimate**	**EFA Component Matrix value**
Idioms	.77	.79
Inferences from Passages	.77	.79
Inferences	.74	.76
Strange Stories	.71	.73
Faux Pas	.66	.69
Why Questions	.64	.69
Deductions	.62	.66
Situational Emotions	.62	.66
Making Predictions	.61	.65
Formulating Solutions	.59	.64
Mental state verbs	.50	.55
Belief-based emotions	.44	.49

It can be seen from the above chart that the CFA and EFA results are in the same loading order and have very similar values. The EFA figures are .02 - .05 more than the CFA values.

The following figure shows the CFA results with sub-section 8 deleted.

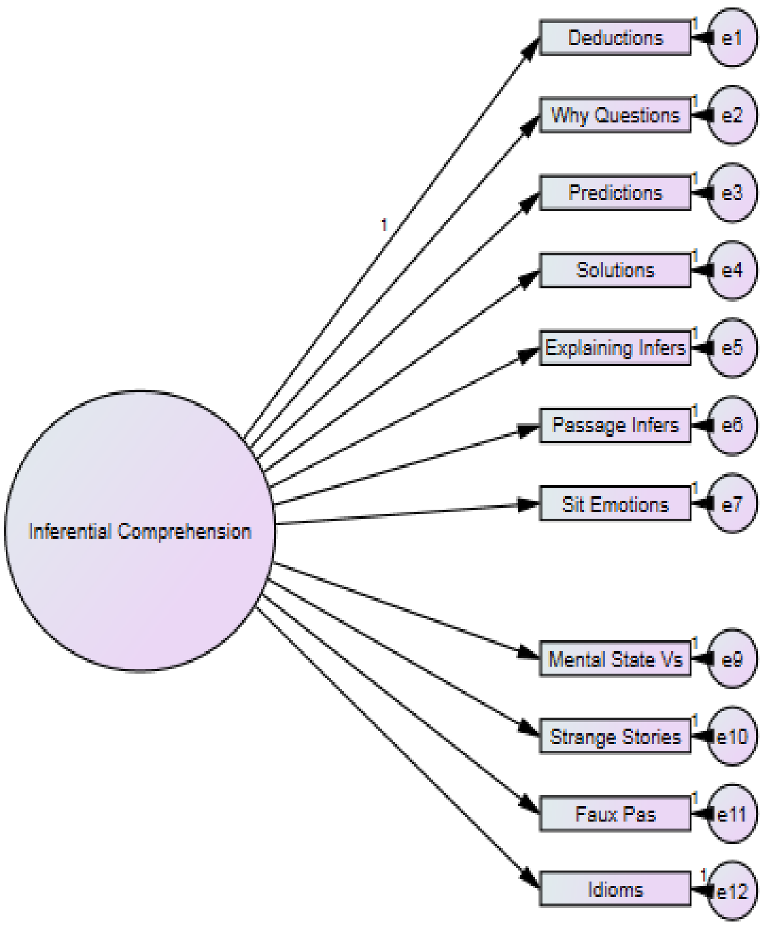


Figure 3.3 Confirmatory factor analysis with sub-test 8 (belief-based emotions) excluded

In this CFA the loadings range from .49 (mental state verbs) to .77 (idioms and passages). The table below demonstrates the loadings in descending order from strongest to weakest.

Table 3.9 CFA standardised estimates values for 11 sub-tests (8 excluded)

Sub-test	**CFA Standardised Estimate**
Idioms	.77
Inferences from Passages	.77
Inferences	.75
Strange Stories	.71
Faux Pas	.66
Why Questions	.64
Making Predictions	.62
Deductions	.61
Situational Emotions	.61
Formulating Solutions	.59
Mental state verbs	.49

The modification index co-variances table demonstrated the following:

Table 3.10 CFA modification index co-variances (<-> indicates co-varying)

	Modification Index (MI)
e11 <--> e12	4.61
e7 <--> e12	10.22
e4 <--> e12	6.82
e3 <--> e11	5.93
e3 <--> e6	5.86
e1 <--> e2	4.04

This looks at any model fit issues in more detail. The largest MI was between e7 and e12 so these were co-varied to see if it improved the model fit (see below).

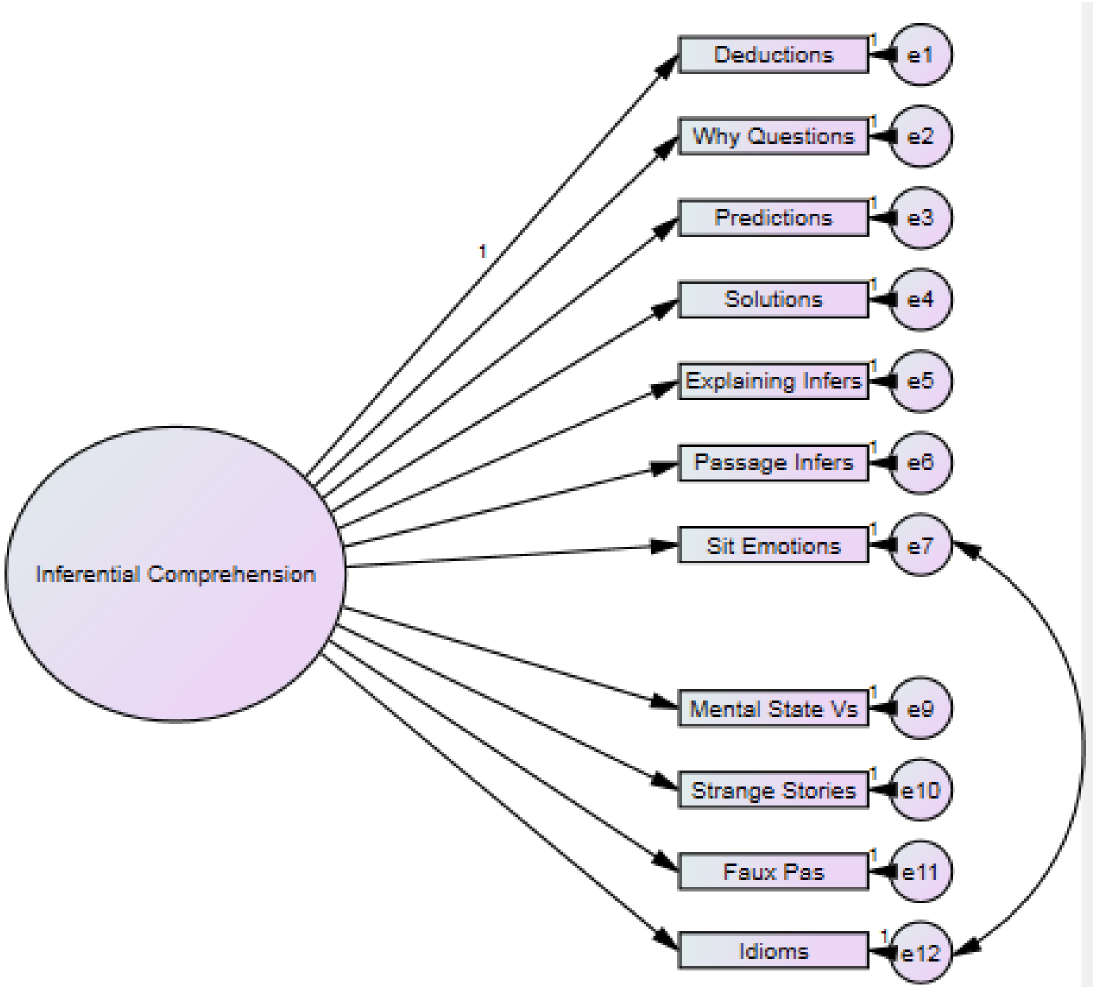


Figure 3.4 CFA with sub-test e8 omitted and with e7 and e12 co-varying

In this CFA the loadings range from .49 (mental state verbs) to .78 (passages). The table below demonstrates the loadings in descending order from strongest to weakest.

Table 3.11 CFA standardised estimates values for 11 sub-tests (8 excluded) and with e7 and e12 co-varying

Sub-test	**CFA Standardised Estimate**
Inferences from Passages	.78
Idioms	.75
Inferences	.75
Strange Stories	.71
Faux Pas	.66
Why Questions	.65
Making Predictions	.62
Deductions	.62
Formulating Solutions	.59
Situational Emotions	.58
Mental state verbs	.49

Model fit measurements

GFI (goodness-of-fit index) reflects the overall discrepancy between the actual correlation covariance matrix versus the predicted covariance matrix (Hatcher, 2013). Values range from 0 to 1 with those nearer to 1 representing a better fit. AGFI (the adjusted goodness-of-fit index) is a parsimonious fit index that rewards less complex models. For both the GFI and the AGFI values exceeding .90 indicate acceptable fit (Hatcher, 2013).

The CFI (comparative fit index) should also be as near to 1.0 as possible. Hatcher (2013) reports that, in the past, an acceptable CFI was >.90 but more recent criteria require it to be > .95

RMSEA (root mean square error of approximation) is another parsimonious fit index. Hatcher (2013) reports that traditionally the fit was seen as acceptable if the RMSEA was <.08, but more recent criteria suggest it should be <.06.

Table 3.12 The HICIT model fit measurements

		CMIN/ DF	P	GFI	AGFI	CFI	RMSEA
Model 1	All 12	1.37	.04	.94	.91	.98	.04
Model 2	e8 omitted	1.477	.02	.94	.91	.98	.05
Model 3	e8 omitted with 7 & 12 co-varied	1.23	.14	.95	.93	.99	.03

The low CMIN/DIF ratios (all below1.5) in all three models indicate a good fit.

P (the probability) can be affected by a large sample size as the test is so sensitive with large samples it can result in an otherwise well-fitting model being rejected (Hatcher, 2013). However, only the third model has a P value which is significant, indicating a poor fit.

The GFIs and AGFIs are over .91 in all three models, indicating a good fit.

The CFI is over .98 in all models, indicating a good fit.

The RMSEA is under .05 for all three models, indicating a good fit.

3.2.3 Test reliability

3.2.3.1 Internal Consistency Reliability: Cronbach's Alpha

The Cronbach's coefficient Alpha (CA) test was carried out on the final results to measure the internal consistency reliability of the sub-tests.

Scale Reliability Results: Individual sections (combined genders) (N=50 in each age group)

Sections 1-8 All age groups included

1) **Deductions: Number of scale items: 8**

Table 3.13 Cronbach's Alpha for sub-section 1 Deductions

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
5-5:11	.49	8
6-6:11	.42	8
7-7:11	.15	7
8-8:11	.37	7
9-9:11	.38	7
CA average across the ages	**.37**	

2) **Why questions: Number of scale items: 12**

Table 3.14 Cronbach's Alpha for sub-section 2 Why Questions

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
5-5:11	.56	12
6-6:11	.49	12
7-7:11	.27	11
8-8:11	.18	12
9-9:11	.25	10
CA average across the ages	**.35**	

3) **Predictions: Number of scale items: 6**

Table 3.15 Cronbach's Alpha for sub-section 3 Prediction

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
5-5:11	.11	6
6-6:11	.09	6
7-7:11	.01	4
8-8:11	-.66	5
9-9:11	-.19	2
CA average across the ages	**.13**	

4) **Formulating solutions. Number of scale items: 12**

Table 3.16 Cronbach's Alpha for sub-section 4 Formulation solutions

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
5-5:11	.61	12
6-6:11	.70	11
7-7:11	.30	11
8-8:11	.56	12
9-9:11	.47	7
CA average across the ages	**.53**	

5) **Making inferences. Number of scale items: 12**

Table 3.17 Cronbach's Alpha for sub-section 5 Making inferences

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
5-5:11	.74	12
6-6:11	.61	12
7-7:11	.52	12
8-8:11	.55	11
9-9:11	.41	11
CA average across the ages	**.57**	

6) **<u>Making inferences from short passages. Number of scale items: 10</u>**

Table 3.18 Cronbach's Alpha for sub-section 6 Making inferences from passages

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
5-5:11	.66	10
6-6:11	.48	10
7-7:11	.57	10
8-8:11	.51	10
9-9:11	.06	8
CA average across the ages	**.46**	

7) **<u>Situation-based emotions. Number of scale items: 16</u>**

Table 3.19 Cronbach's Alpha for sub-section 7 Situation-based emotions

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
5-5:11	.54	16
6-6:11	.44	13
7-7:11	.52	12
8-8:11	.40	13
9-9:11	.42	9
CA average across the ages	**.46**	

8) **<u>Belief-based emotions. Number of scale items: 8</u>**

Table 3.20 Cronbach's Alpha for sub-section 8 Belief-based emotions

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
5-5:11	.29	8
6-6:11	.27	8
7-7:11	.42	7
8-8:11	.51	8
9-9:11	.38	7
CA average across the ages	**.37**	

Sections 9-12. Age groups 6:00-9:11 only

9) **Mental state verbs. Number of scale items: 12**

Table 3.21 Cronbach's Alpha for sub-section 9 Mental state verbs

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
6-6:11	.58	12
7-7:11	.67	12
8-8:11	.55	12
9-9:11	.45	12
CA average across the ages	**.56**	

10) **Strange Stories. Number of scale items: 12**

Table 3.22 Cronbach's Alpha for sub-section 10 Strange Stories

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
6-6:11	.67	12
7-7:11	.62	11
8-8:11	.30	11
9-9:11	.55	10
CA average across the ages	**.54**	

11) **Faux pas. Number of scale items: 12**

Table 3.23 Cronbach's Alpha for sub-section 11 Faux pas

Age Group	**Cronbach's Alpha (CA)**	**Number of items included in the analysis**
6-6:11	.67	12
7-7:11	.58	12
8-8:11	.52	11
9-9:11	.58	11
CA average across the ages	**.59**	

12) **Idioms. Number of scale items: 20**

Table 3.24 Cronbach's Alpha for sub-section 12 Idioms

Age Group	Cronbach's Alpha (CA)	Number of items included in the analysis
6-6:11	.55	17
7-7:11	.68	20
8-8:11	.73	20
9-9:11	.62	20
CA average across the ages	**.65**	

Where the number of items analysed is less than the total number of section items this is because one or more question is answered either totally correctly or totally incorrectly by all children in that age group

McCauley (2001) states that a Cronbach Alpha value of less than .2 is a slight, negligible correlation, .2-.4 is a low correlation, .4-.7 is a moderate correlation, .7-.9 is a high correlation, and over .9 is a very high correlation.

It can be seen from the above results that the CA values for the HICIT sub-tests range from .13 to .65, indicating negligible to moderate reliability. The CA averages across the ages for the sub-tests in order of highest to lowest reliability are: Idioms .65, Faux Pas .59, Making inferences .57, Mental state verbs .56, Strange Stories .54, Formulating solutions .53, Making inferences from short passages .46, Situation-based emotions .46, Deductions .37, Belief-based emotions .37, Why questions .35, Predictions .13.

3.2.3.2 Internal Consistency Reliability: Rasch Analysis

Rasch Analysis is a form of Item Response Theory (IRT) which measures the degree to which the test scoring is justified by the data collected, ie that the total score means the same across the compared groups. It was discussed in detail in the previous methodology chapter. The HICIT assessment produces dichotomous data (responses are scored 0 or 1) so the data were analysed using the dichotomous Rasch model. The results are represented in figures 3.5 to 3.7 below. Figure 3.5 represents test item difficulty across all 12 sections for children aged 6:00-9:11. It includes outliers. Figure 3.6 is the same but excludes the outliers.

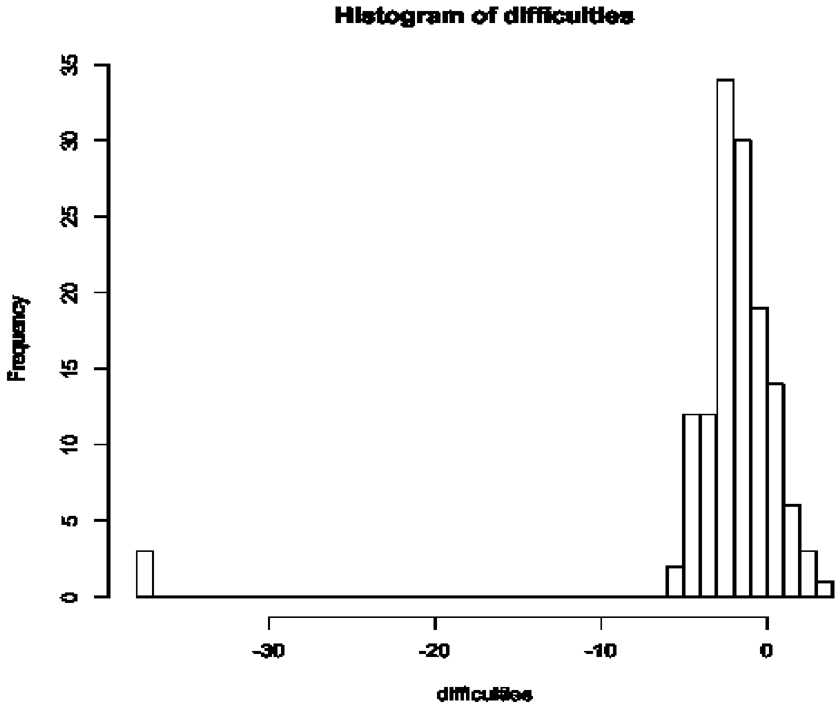


Figure 3.5 Histogram of difficulties across all 12 sections (ages 6:00-9:11) including outliers: Situational Emotions questions 4 (happy), 5 (happy), 14 (happy)

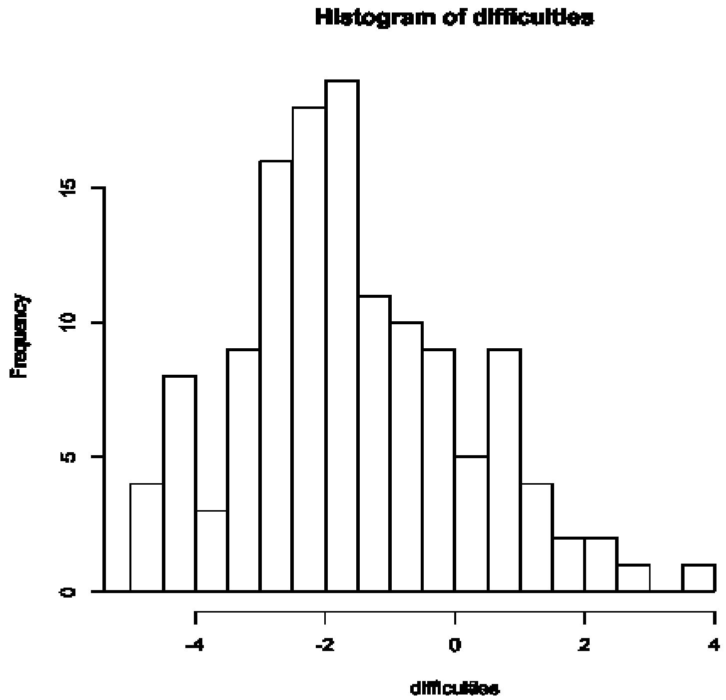


Figure 3.6 Histogram of difficulties across all 12 sections (ages 6:00-9:11) excluding outliers.

<u>Comparison of Rasch Analysis total scores and SPSS analysis total scores</u>

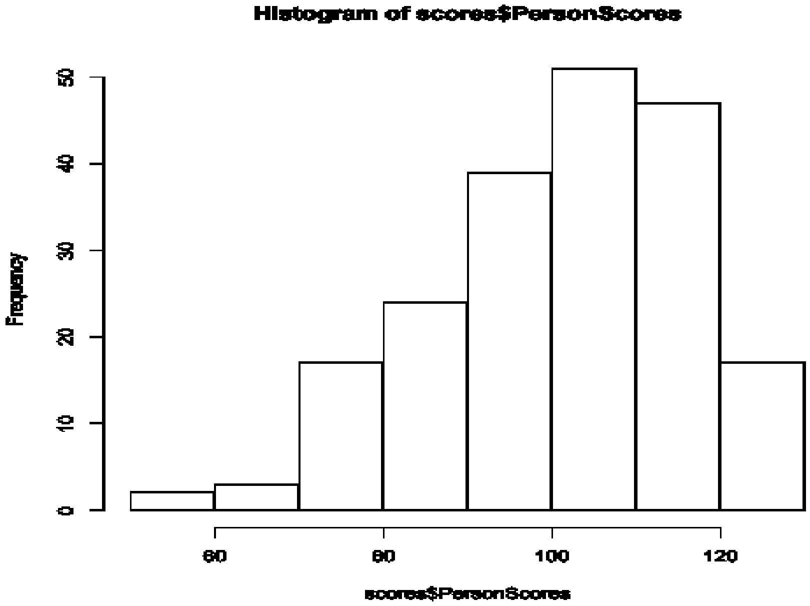


Figure 3.7 Rasch Analysis histogram of total test scores (6:00-9:11 year-olds)

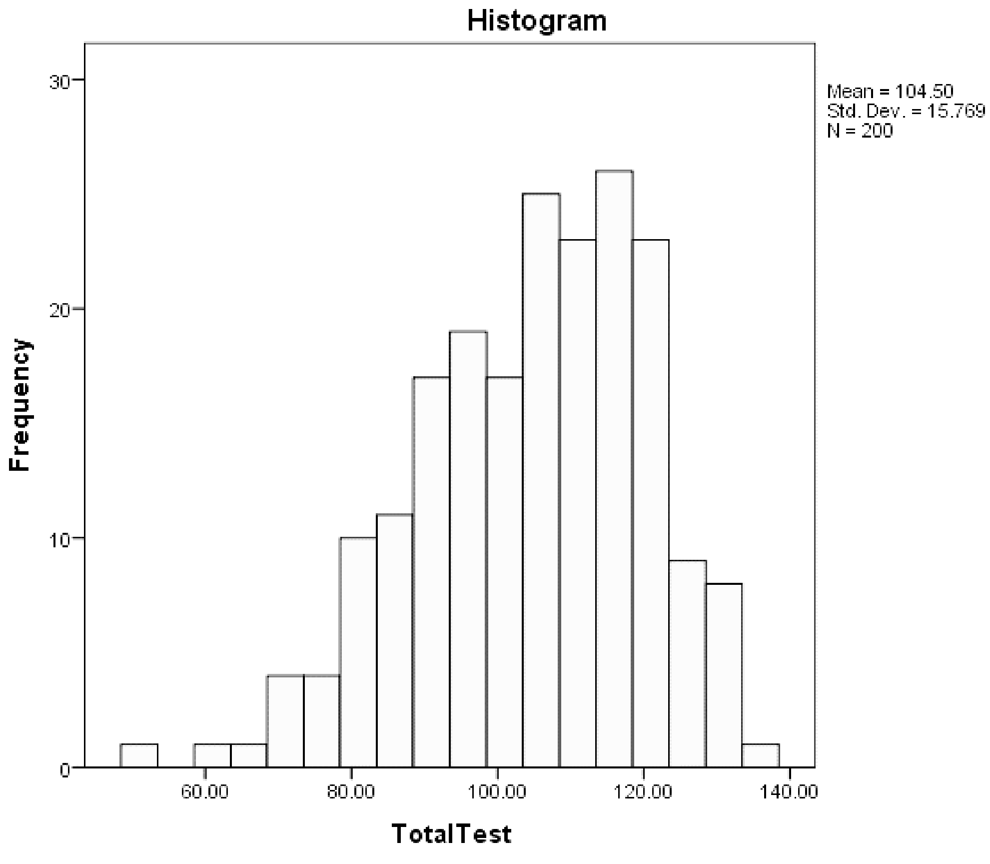


Figure 3.8 SPSS histogram of total test scores (6:00-9:11 year-olds)

Comparison of the Rasch analysis and the SPSS total results (6:00-9:11 year-olds) demonstrates very similar shaped histograms. Both are negatively skewed (ie there is a pile up of cases to the right and the left tail longer).

3.2.3.3 Inter-rater reliability

Researcher to supervisor inter-rater reliability

The researcher recorded 10% of the children she assessed so her PhD supervisors could moderate her marking. 10% of the 200 final study children is 20.

Four children from each age group (5-5:11, 6-6:11, 7-7:11, 8-8:11 and 9-9:11) were audio-recorded: a total of 20 children. The supervisors listened to each audio-recording and marked it using the marking criteria and a blank assessment form so that they were not aware of the researcher's scoring. The first two children were moderated independently by both supervisors so any errors or ambiguities in the assessment form and marking criteria could be discussed before the rest of the sample was moderated. No significant errors or ambiguities were detected.

Table 3.25 Researcher to supervisor inter-rater reliability scores for HICIT marking

Results table code: Green is supervisor 1, blue is supervisor 2

Supervisor	Child's age	Supervisor's score /140	Researcher's score/140	Number of Qs marked differently	% Concordance
1	5:10	57	54	3	97.9%
1	5:10	64	61	3	97.9%
1	5:10	86	87	1	99.3%
1	5:11	73	72	1	99.3%
1	7:03	105	108	3	97.9%
1	7:04	100	101	1	99.3%
1	8.00	88	90	2	98.6%
1	8.05	123	126	3	97.9%
1	9:00	116	116	0	100%
1	9:05	123	120	3	97.9%
2	8:00	91	90	1	99.3%
2	8:05	126	126	0	100%
2	9:09	120	120	2	98.6%

The overall supervisor-researcher inter-rater reliability concordance rate for 13 participants is: 98.79%.

Student to researcher

The eight volunteer MMU students were asked to mark 1 of their assessments using the marking criteria sent to them. The researcher then marked these herself to moderate the students' scoring. The student assessing the 5 year-olds did not mark a completed assessment as only the first 8 sections were carried out with this age group. Consequently, only 7 students were moderated. The percentage concordance is calculated by dividing the total agreed marks by the total agreed plus disagreed marks and multiplying by 100.

The results are displayed below:

Table 3.26 Researcher to student inter-rater reliability scores for HICIT marking

Student no.	**Child's age**	**Student's score**	**Researcher's score**	**Number of Qs marked differently**	**% Concordance**
1	6:00	98/140	97/140	1	99.3%
2	7:09	107/140	103/140	4	97.14%
3	7:10	109/140	110/140	1	99.3%
4	8:00	115/140	108/140	7	95%
5	8:09	74/140	74/140	0	100%
6	9:02	109/140	109/140	2	98.6%
7	9:09	122/140	122/140	0	100%

Student number 4 did not adhere fully to the marking criteria.

For example:

Sub- Section 5 - Explaining Inferences (How do you know?)

1. How do you know that someone is tired? The student accepted 'sleep in their eyes'.
2. 5. How do you know that someone is too hot? The student accepted 'by their face'

8. How do you know that someone has got toothache? The student accepted 'Mum would phone us. Wobbly tooth'

10. How do you know what's inside a box of something at the supermarket? The student accepted 'Shake it – sounds'

11. How do you know that someone thinks there is a bad smell in the room? The student accepted 'Smell it – comes up your nose'

10) Strange Stories

Pretend

John and Rashid are playing spacemen. John picks up a bowl and puts it on his head. He says: ‘This bowl is a space helmet.’

1) Is it true what John says?

2) Why does he say it? The student accepted ‘He just wants to be silly’.

12) Idioms

9. She got out of the wrong side of bed today. The student accepted ‘sad’.

The overall student-researcher inter-rater reliability concordance rate if all 7 students are used is 98.5%.

If student 4 is removed from this calculation due to her not sticking to the marking criteria then the rate is 99.1%

The overall researcher to supervisor/student inter-rater reliability concordance rate is: 98.63% if student 4 is included or 98.93% if her results are excluded.

Section 3.3 Final study results

3.3.1 Summary of final participants

Table 3.26 below summarises the 250 participants. These were described in detail in section 2.4.6.1 of the previous methodology chapter (chapter 2). The researcher carried out the fifty pilot study assessments and 122 final assessments and eight student volunteers carried out the remaining 78 assessments.

(S) = carried out by student

(R) = carried out by Researcher

(P) (R) = Pilot study carried out by the Researcher

Table 3.27 Summary of the final HICIT study participants

Area of GB	**Number of children assessed**	**Age Group assessed**	**Percentage Free School Meals**
North West (S)	10	5-5:11	10.9%
North Wales (S)	10	6-6:11	27%
North West (S)	3	6-6:11	7%
East Midlands (S)	1	7-7:11	4.8%
North West (S)	14	7-7:11	9.8%
North West (S)	10	7-7:11	50.3%
West Yorkshire (S)	3	8-8:11	6%
East Midlands (S)	10	8-8:11	4.8%
West Midlands (S)	8	8-8:11	7%
West Yorkshire (S)	6	9-9:11	6%
West Cheshire (S)	7	9-9:11	9.6%
North West (R)	72	5-5:11 6-6:11 7-7:11 8-8:11 9-9:11	9.6%
North West (R)	46	5-5:11 6-6:11 7-7:11 8-8:11 9-9:11	53.6%
North West (P) (R)	30	5-5:11 6-6:11 7-7:11	0%
North West (P) (R)	20	8-8:11 9-9:11	9%

<u>Average Percentage of Free School Meals (FSM):</u>

This was calculated by multiplying each number of children by the corresponding percentage FSM, adding each value together and then dividing by 250 (total number of children assessed).

The overall mean FSM percentage was 21.4%, slightly above the national average of 19.3%. This indicates that a good balance of schools across a range of socio-economic areas was used.

3.3.2 Descriptive statistics

3.3.2.1 Final Total Scores: 5:00-5:11 year-olds

As noted in the previous methodology chapter, only sections 1-8 of the final HICIT were completed with this age group, as sections 9-12 were too difficult for this age group in the pilot study.

The total maximum possible score from these 8 sub-sections is 84.

The Kolmogorov-Smirnov test demonstrated that the total test data are normally distributed. The mean score was therefore selected as the measure of central tendency as this is a parametric measurement.

Table 3.28 Total HICIT scores for 5:00-5:11 year-olds

Age Band	Sex	Mean (/84)	Range
5:00-5:11	Male	48.12	32-64
5:00-5:11	Female	48.48	27-67
5:00-5:11	Combined	48.3	27-67

3.3.2.2 Final Total Scores: 6:00 - 9:11 year-olds

All 12 sections completed with these age groups. The total maximum score is 140.

Scores by gender

Table 3.29 Total HICIT scores for 6:00-9:11 year-olds separate genders

Age Band	Sex	Mean (/140)	Range
6:00-6:11	Male	89.48	72-123
6:00-6:11	Female	86.72	51-108
7:00-7:11	Male	99.24	81-120
7:00-7:11	Female	106.36	80-118
8.00-8:11	Male	105.72	90-122
8:00-8:11	Female	108.28	85-126
9:00-9:11	Male	118.8	100-132
9:00-9:11	Female	120.44	100-134

Total combined scores

Table 3.30 Total HICIT scores for 6:00-9:11 year-olds genders combined

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig.
6-6:11	88.14	88.36	89	190.25	13.79	51-123 (72)	20	-.167	.165	.2
7-7:11	102.9	103.13	104.5	103.24	10.16	80-120 (40)	17.25	-.35	-.802	.2
8-8:11	107.06	107.27	109	121.04	11.00	85-126 (41)	19.25	-.26	-1.025	.2
9-9:11	119.9	120.24	120	70.91	8.42	100-134 (34)	13.25	-.45	-.244	.2

*The Kolmogorov-Smirnov test of normality (**Sig.** value of more than .05 indicates a normal distribution, .2 is a lower bound of the true significance).

The Kolmogorov-Smirnov Significance value of .2 for each age group indicates that there is a normal distribution within each of these groups (a value of more than .05 is significant).

These normative data demonstrate that the mean total test score increases gradually as the children get older, indicating that children's inferential and idiom comprehension, as measured by the final HICIT, improves with age.

Histogram and scattergram examples from all the age groups

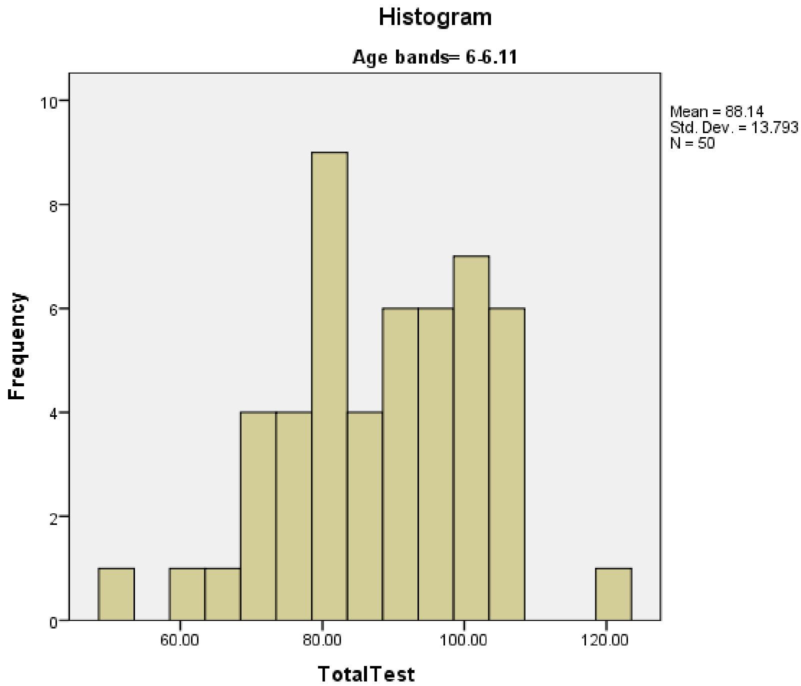


Figure 3.9 Total test scores histogram for 6:00-6:11 year-olds

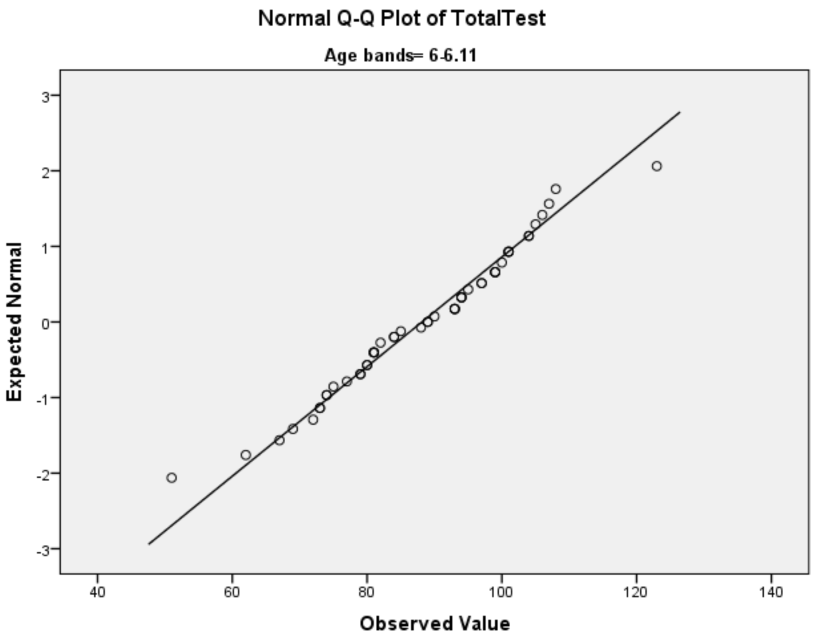


Figure 3.10 Scattergram of total scores for 6:00-6:11 year-olds

There are two 'extreme outliers' in the 6:00 to 6:11 age group, one a very high scorer and one a very low scorer: A boy aged 6:00 scored 123/140 and a girl aged 6:06 scored 51/140. If these 2 outliers are taken out the results become:
Boys: Mean 88.3 - Range 72-107
Girls: Mean 88.2 - Range 62-108
Combined: Mean 88.25 - Range 62-108

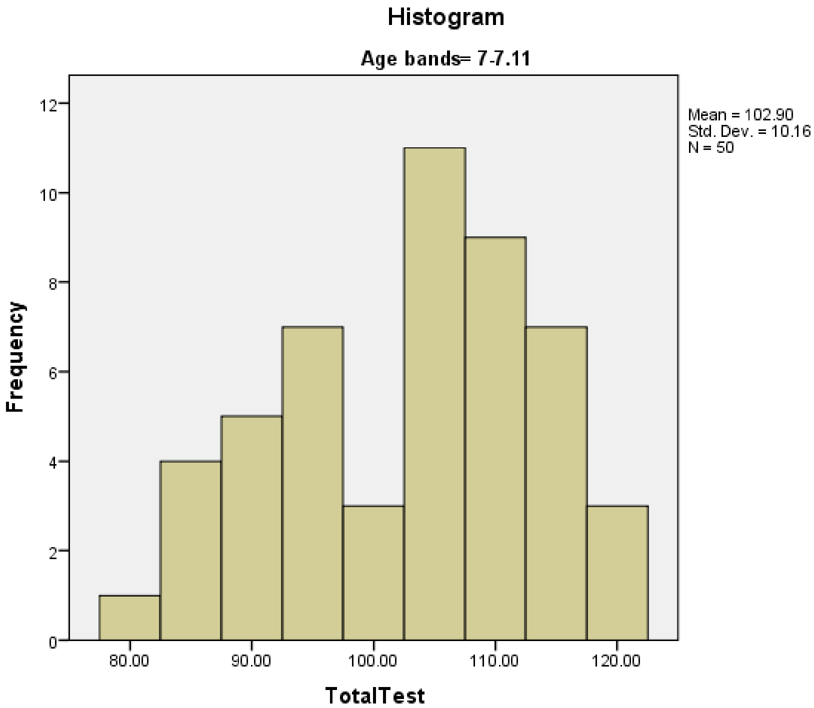


Figure 3.11 Total test scores histogram for 7:00-7:11 year-olds

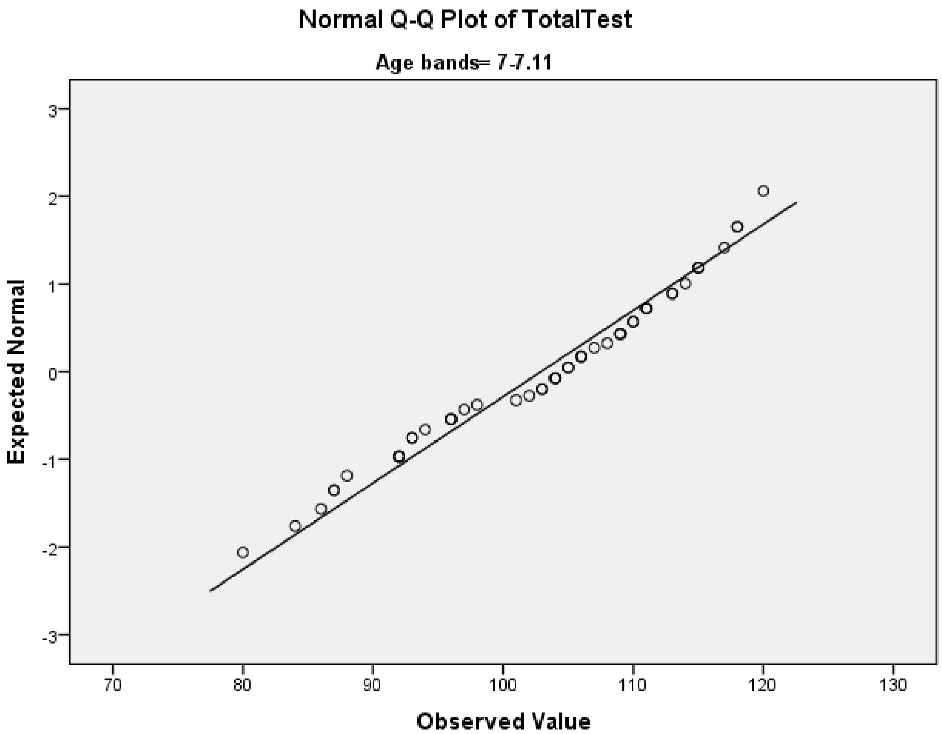


Figure 3.12 Scattergram of total scores for 7:00-7:11 year-olds

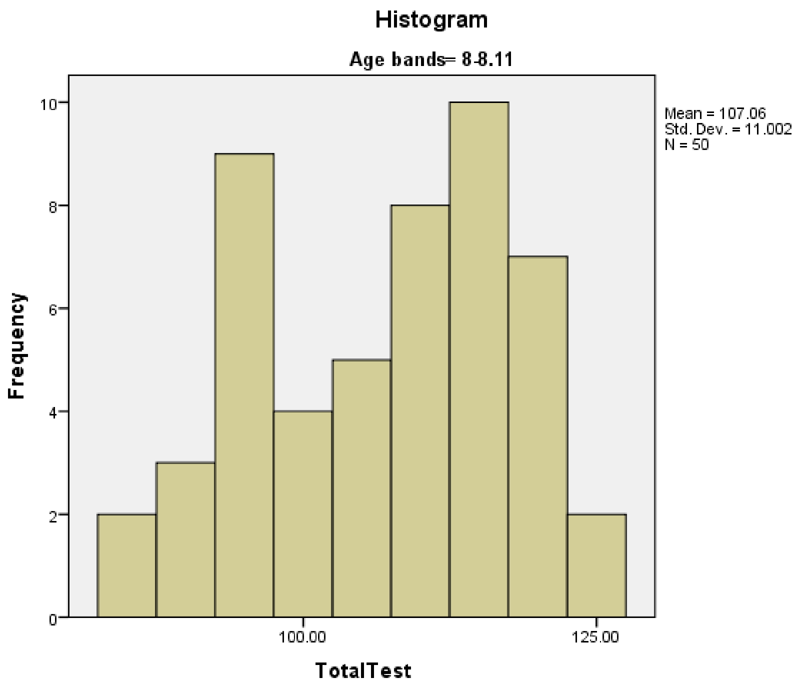


Figure 3.13 Total test scores histogram for 8:00-8:11 year-olds

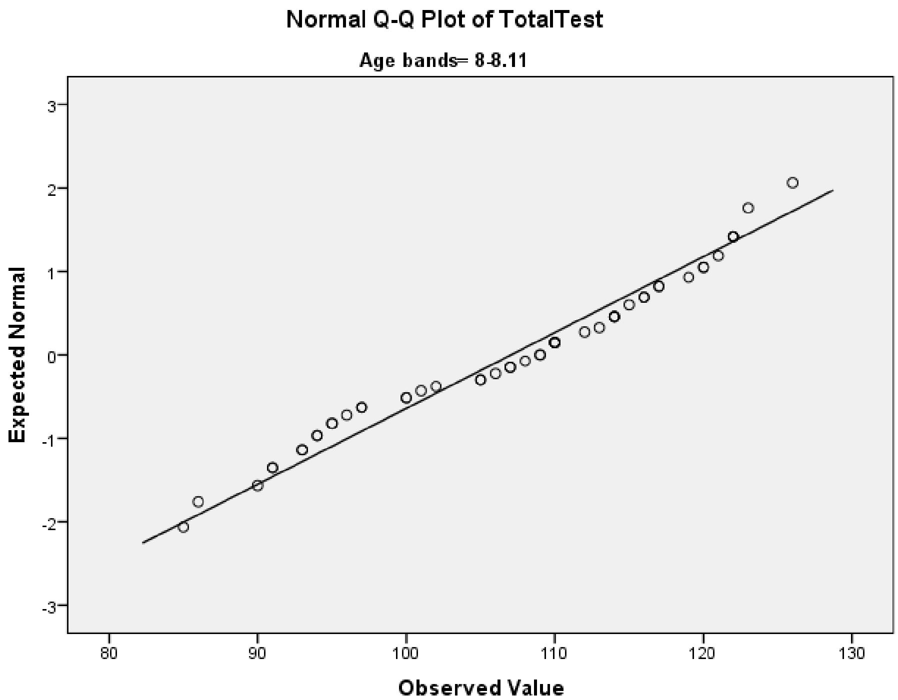


Figure 3.14 Scattergram of total scores for 8:00-8:11 year-olds

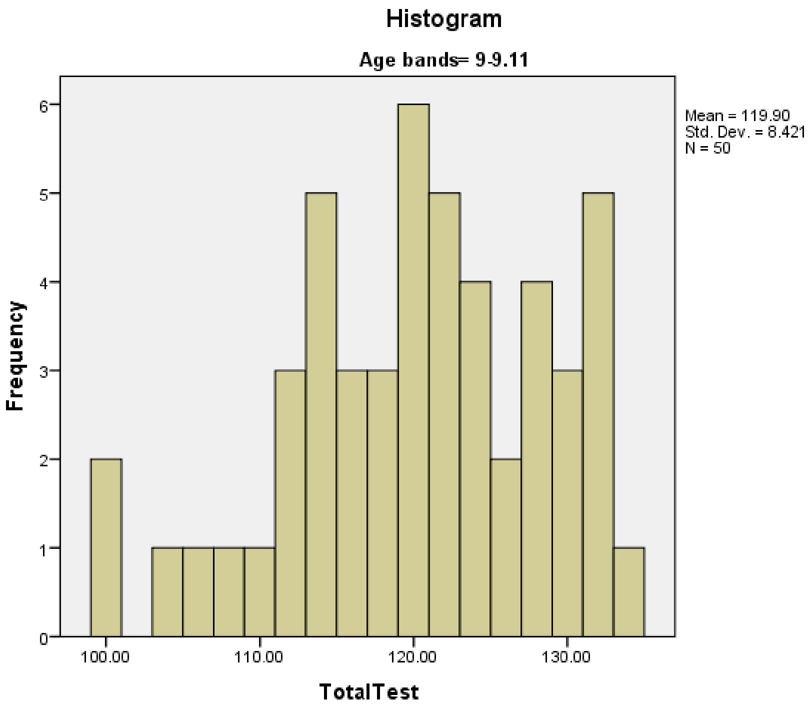


Figure 3.15 Total test scores histogram for 9:00-9:11 year-olds

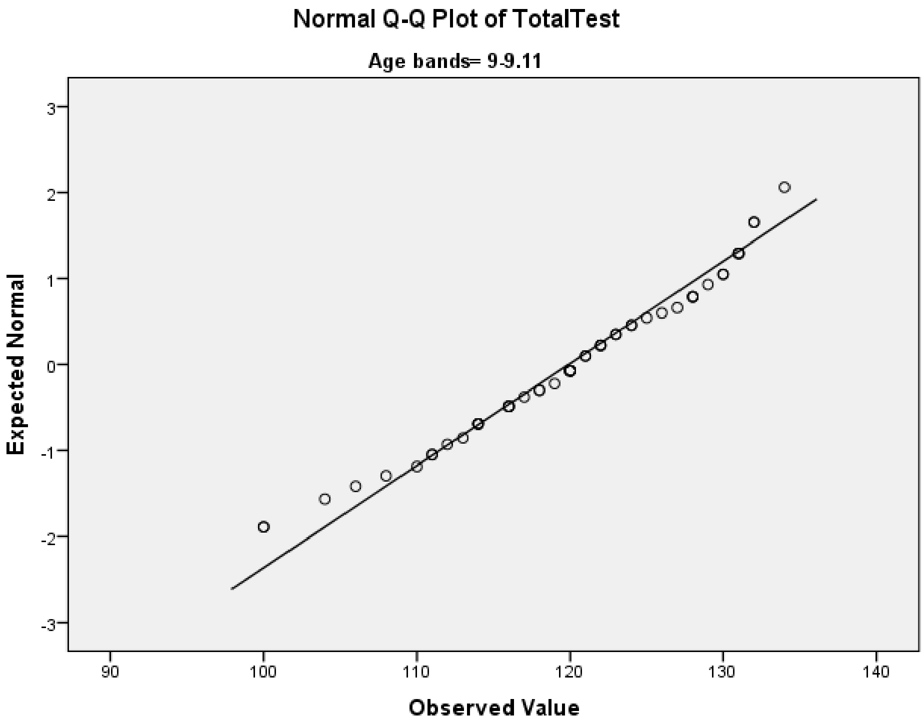


Figure 3.16 Scattergram of total scores for 9:00-9:11 year-olds

Individual section results (combined genders)
N=50 in all age groups
Sections 1-8: all age groups are included, sections 9-12: only 6:00-9:11 age groups are included.

1) Deductions

In the first sub-section of the HICIT the child has to deduce what an object is from a series of clues given verbally.

The maximum possible score is 8

Table 3.31 Descriptive statistics for section 1

Age	Mean	5% Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig.
5-5:11	4.48	4.5	4.5	2.87	1.70	1-7 (6)	3	-.12	-1.148	.00
6-6:11	5.48	5.48	5.5	2.21	1.49	3-8 (5)	2.25	0.79	-.803	.02
7-7:11	6.44	6.48	7	1.31	1.15	3-8 (5)	1.25	-.53	.058	.00
8-8:11	6.72	6.79	7	1.31	1.14	4-8 (4)	2	-.61	-.424	.00
9-9:11	7.06	7.17	7	1.14	1.07	3-8 (5)	2	-1.41	2.906	.00

*The Kolmogorov-Smirnov test of normality (**Sig.** value of more than .05 indicates a normal distribution)

There are no normal distributions in this section.
The age group 6-6:11 has the nearest to a normal distribution is slightly positively skewed (ie there is a pile up of cases to the left and the right tail is too long).
All of the others are negatively skewed (ie there is a pile up of cases to the right and the left tail is too long).

2) Why questions

In the second sub-section of the HICIT, the child has to answer a range of different why questions. The maximum possible score is 12

Table 3.32 Descriptive statistics for section 2

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
5-5:11	6.6	6.62	6.5	5.43	2.33	0-11 (11)	3	-.25	.16	.20
6-6:11	8.64	8.69	9	3.91	2.0	4-12 (8)	3	-.58	-.41	.00
7-7:11	10.02	10.1	10	1.82	1.35	6-12 (6)	2	-.82	.76	.00
8-8:11	10.32	10.4	10	1.53	1.24	6-12 (6)	1	-.99	1.92	.00
9-9:11	11.04	11.12	11	.79	.89	8-12 (4)	1	-1.2	2.1	.00

There is one normal distribution in this section – age group 5:00 to 5.11 (with a negative skew). All of the others are negatively skewed.

3) Predictions

In the third sub-section of the HICIT, the child has to answer 'What would happen if ….?' questions. The maximum possible score is 6

Table 3.33 Descriptive statistics for section 3

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
5-5:11	4.1	4.16	4	1.32	1.15	1-6 (5)	2	-.54	.70	.00
6-6:11	5.06	5.12	5	.79	.89	3-6 (3)	1.25	-.66	-.27	.00
7-7:11	5.58	5.64	6	.37	.61	4-6 (2)	1	-1.17	.39	.00
8-8:11	5.56	5.57	6	.25	.50	5-6 (1)	1	-.25	-2.02	.00
9-9:11	5.84	5.88	6	.14	.37	5-6 (1)	0	-1.88	1.60	.00

There are no normal distributions in this section and all are negatively skewed

4) <u>Formulating solutions</u>

In the fourth sub-section of the HICIT the child has to answer 'What could you do if …?' questions.

The maximum possible score is 12

Table 3.34 Descriptive statistics for section 4

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
5-5:11	8.6	8.7	9	4.74	2.18	3-12 (9)	3.25	-.64	-.091	.01
6-6:11	10.22	10.46	11	3.73	1.93	1-12 (11)	1.25	-2.56	9.856	.00
7-7:11	11.12	11.21	11	1.05	1.02	8-12 (4)	1.25	-1.08	.635	.00
8-8:11	11.14	11.31	12	1.63	1.28	7-12 (5)	1	-1.86	3.294	.00
9-9:11	11.51	11.65	12	.76	.87	8-12 (4)	1	-2.42	6.493	.00

There are no normal distributions in this section and all are negatively skewed.

5) <u>Making inferences</u>

In the fifth sub-section of the HICIT the child has to answer 'How do you know that ….? questions.

The maximum possible score is 12

Table 3.35 Descriptive statistics for section 5

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
5-5:11	5.1	5.11	5	7.80	2.79	0-10 (10)	4.25	.045	-1.04	.14
6-6:11	8.14	8.19	9	5.23	2.29	3-12 (9)	3.25	-.403	-.73	.00
7-7:11	9.06	9.11	9	3.69	1.92	5-12 (7)	3	-.251	-.77	.00
8-8:11	9.92	9.99	10	3.06	1.75	6-12 (6)	3	-.374	-.86	.00
9-9:11	11.16	11.25	11	.81	.90	8-12 (4)	1	-1.24	2.06	.00

There is one normal distribution in this section – age group 5:00 to 5.11 (with a positive skew). All of the others are negatively skewed.

6) <u>Making inferences from short passages</u>

In the sixth sub-section of the HICIT the child answers inference questions about a short paragraph.

The maximum possible score is 10

Table 3.36 Descriptive statistics for section 6

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
5-5:11	4.66	4.6778	5	4.882	2.20953	0-9 (9)	4	-.175	-.784	.095
6-6:11	7.04	7.0556	7	3.141	1.77235	4-10 (6)	3	-.292	-1.012	.001
7-7:11	8.52	8.6222	9	2.418	1.55498	4-10 (6)	2	-.870	-.004	.000
8-8:11	8.72	8.8778	9	2.042	1.42914	4-10 (6)	2	-1.535	2.384	.000
9-9:11	9.3265	9.4082	9	.641	.80072	7-10 (3)	1	-1.178	1.175	.000

There is one normal distribution in this section – age group 5:00 to 5.11 (with a negative skew). All of the others are negatively skewed.

7) <u>Situation-based emotions</u>

In the seventh sub-section of the HICIT the child has to decide which emotion out of a choice of 4 best fits different scenarios.

The maximum possible score is 16

Table 3.37 Descriptive statistics for section 7

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
5-5:11	9.64	9.68	10	5.05	2.25	3-14 (11)	3	-.30	.55	.02
6-6:11	11.76	11.74	12	3.57	1.90	8-16 (8)	3	.021	-.85	.01
7-7:11	13.26	13.37	14	3.58	1.90	8-16 (8)	2.25	-.94	.60	.00
8-8:11	13.62	13.68	14	2.73	1.65	10-16 (6)	2	-.40	-.51	.01
9-9:11	14.78	14.85	15	1.51	1.23	12-16 (4)	2	-.67	-.70	.00

There are no normal distributions in this section.

Age group 6-6:11 is positively skewed. All other age groups are negatively skewed.

8) <u>Belief-based emotions</u>

In the eighth sub-section of the HICIT the child has to decide if belief-based scenarios would make someone happy or sad.

The maximum possible score is 8

Table 3.38 Descriptive statistics for section 8

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
5-5:11	5.18	5.2	5	1.82	1.35	2-8 (6)	2	-.08	-.56	.01
6-6:11	5.48	5.48	5	1.81	1.34	3-8 (5)	1.	.26	-.45	.00
7-7:11	6.1	6.16	6	1.89	1.37	2-8 (6)	2	-.53	.10	.00
8-8:11	6.34	6.44	6	2.15	1.47	2-8 (6)	2	-.79	.43	.00
9-9:11	6.94	7.01	7	1.27	1.13	4-8 (4)	2	-.70	-.54	.00

There are no normal distributions in this section. They are all negatively skewed.

Sections 9-12. Age groups from 6 only

9) Mental state verbs

In the ninth sub-section of the HICIT the child has to decide if a mental state verb means that something definitely does or doesn't happen, or might happen (choice of yes/ no/ maybe).

The maximum possible score is 12

Table 3.39 Descriptive statistics for section 9

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
6-6:11	7.86	7.89	8	4.61	2.15	3-12 (9)	3	-.3	-.17	.00
7-7:11	8.68	8.72	9	4.55	2.13	4-12 (8)	3.25	-.25	-.82	.00
8-8:11	9	9.08	9	3.80	1.95	4-12 (8)	3	-.62	-.13	.02
9-9:11	9.96	10.03	10	2.42	1.55	6-12 (6)	2	-.79	-.15	.00

There are no normal distributions in this section. They are all negatively skewed.

10) Strange Stories

In the tenth sub-section of the HICIT the child has to work out what other people are thinking.

The maximum possible score is 12

Table 3.40 Descriptive statistics for section 10

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
6-6:11	8.1	8.14	8	4.95	2.22	3-12 (9)	3	-.46	-.34	.01
7-7:11	9.9	10	10	3.31	1.82	5-12 (7)	2	-.97	.48	.00
8-8:11	9.8	9.9	10	2.06	1.43	6-12 (6)	2	-.49	-.07	.00
9-9:11	10.84	11	11	1.93	1.39	6-12 (6)	2	-1.64	2.85	.00

There are no normal distributions in this section. They are all negatively skewed

11) Faux pas

In the eleventh sub-section of the HICIT the child has to identify faux pas and work out what other people are feeling.

The maximum possible score is 12

Table 3.41 Descriptive statistics for section 11

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
6-6:11	7.2	7.16	7	3.92	1.98	3-12 (9)	2	.30	.9	.00
7-7:11	8.54	8.56	8	3.27	1.81	4-12 (8)	2.25	-.02	-.12	.00
8-8:11	8.84	8.84	9	2.75	1.66	5-12 (7)	2	.1	-.31	.01
9-9:11	10.08	10.19	10	2.79	1.67	6-12 (6)	3	-.58	-.34	.00

There are no normal distributions in this section. Age groups 6:00 to 6:11 and 8:00 to 8:11 are positively skewed and the other two are negatively skewed.

12) Idioms

In the twelfth sub-section of the HICIT the child has to define 20 common idioms. The maximum possible score is 20

Table 3.42 Descriptive statistics for section 12

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig
6-6:11	3.16	3.03	3	4.46	2.11	0-10 (10)	2	.92	1.22	.00
7-7:11	5.72	5.77	5	9.43	3.07	0-11 (11)	5.25	.01	-1.07	.15
8-8:11	7.04	7.09	7	11.88	3.45	0-13 (13)	5	.05	-.74	.00
9-9:11	11.4	11.42	12	9.14	3.02	5-18 (13)	3.25	-.09	-.26	.20

There are 2 normal distributions in this section: age groups 7:00 to 7:11 and 9:00 to 9:11. All are positively skewed, apart from 9-9:11, which is negatively skewed.

3.3.3 Inferential statistics

The Kolmogorov-Smirnov test demonstrated that the total test data are normally distributed. The parametric tests one-way and two-way ANOVAs could therefore be applied to the results to look at the interaction between socio-economic factors and total test score, and between age, gender and total test score.

A one-way ANOVA was carried out to explore the effect of socio-economic status on the total test score.

3.3.3.1 Socio-economic status and total test results

Table 3.43 One-way ANOVA results. The effect of socio-economic status on test scores

% Free School Meals	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
.00	20	96.85	13.56	3.03	90.50	103.20	51.00	113.00
4.80	10	107.70	11.00	3.47	99.85	115.56	86.00	122.00
6.00	9	115.78	4.66	1.55	112.20	119.36	110.00	122.00
7.00	12	103.08	15.64	4.52	93.15	113.02	80.00	132.00
9.00	19	118.37	8.14	1.87	114.45	122.30	97.00	131.00
9.60	60	107.95	16.55	2.14	103.67	112.23	69.00	134.00
9.80	14	106.71	10.31	2.75	100.76	112.67	84.00	120.00
27.00	10	89.80	18.07	5.72	76.87	102.73	62.00	123.00
50.30	10	98.00	10.98	3.47	90.14	105.86	80.00	115.00
53.60	36	97.47	15.15	2.53	92.35	102.60	67.00	132.00
Total	200	104.50	15.77	1.12	102.30	106.70	51.00	134.00

Table 3.44 Test of homogeneity of variances

Levene Statistic	df1	df2	Sig.
3.03	9	190	.00

Table 3.45 Robust tests of equality of means

	Statistic[a]	df1	df2	Sig.
Welch	10.59	9	51.43	.00
Brown-Forsythe	7.52	9	99.23	.00

a. Asymptotically F distributed.

The Levene test indicates that the homogeneity of variance assumption has been violated as the significance (Sig). is **less than** 0.05. This means that more robust tests of equality of means than the ANOVA need to be used. These are the Welch and Brown-Forsythe tests above. The Sig. values here are both .000. To be significant these should be less than or equal to .05.

The above results indicate that there is a significant difference in total test scores across the socio-economic groups.

This could confirm the alternate hypothesis (Ha) Hypothesis 3 (that socio-economic status has an effect on inferential and idiom comprehension). However, the age groups differed in each socio-economic group so more detailed analysis of the results in the form of a post hoc test was required. These results will be analysed further in the final discussion and conclusion chapter.

Post hoc tests

Post hoc tests should only be carried out if there is a significant main effect in the overall analysis of variance test. A significant main effect for socio-economic status affecting test scores was obtained, so post hoc tests could be applied.

Post hoc tests for socio-economic status

As different age bands were assessed in the different socio-economic groups the only cohorts that can be compared are:

1) **Cohort 1: Socio-economic status**

Table 3.46 Cohort 1. The 2 groups that are comparable for socio-economic status

Region	Number of children	Age Range of children	Free School Meal (FSM) %	Mean Total Test Score
North West (R) and West Cheshire (S)	60	6:00-9:11	9.6%	107.96
North West (R)	36	6:00-9:11	53.6%	97.47

The breakdown of the ages of the children in each age band in the above cohort is:
Table 3.47 Breakdown of number of children of different ages in each group

Age Band	No of children in age band (9.6% FSM)	No of children in age band (53.6% FSM)
6-6:11	13	14
7-7:11	13	2
8-8:11	6	14
9-9:11	28	6

The mean age of the children in the 9.6% FSM cohort was 8:04 years.
The mean age of the children in the 53.6% FSM cohort was 7:10 years (6 months less than the 9.6% FSM cohort).
An independent samples T test was carried out on this data.
The results are shown below:

The total number of 53.6% FSM children is 45 for the individual sub-tests 1-8 but is 36 for the individual sub-tests 9-12 and the Total test score.

The colour highlighting is explained in the total test score analysis section below.

Table 3.48 Cohort 1 group statistics

	FSM %age	N	Mean	Std. Deviation	Std. Error Mean
TotalTest	9.6	60	107.95	16.55	2.14
	53.6	36	97.47	15.15	2.53

Table 3.49 Cohort 1 Independent samples T test

		Levene's Test		t-test for Equality of Means						
									95% CID	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Total Test	EVA	1.02	.32	3.097	94	.00	10.48	3.38	3.76	17.19
	EVNA			3.17	79.04	.00	10.48	3.31	3.90	17.06

EVA stands for Equal variances assumed; EVNA stands for Equal variances NOT assumed

CID stands for Confidence Interval of the Difference.

Total Test Score analysis

The significance value for Levene’s Test for equality of variances is .32 (highlighted in yellow above). As this is larger than .05 equal variances can be assumed, so the first line of the table was used for the T test result.

The green highlighted T test result (.00) is less than .05, indicating that the difference in means between the two cohorts is statistically significant.

There is a significant difference between the scores for the 9.6% FSM cohort (M =107.95, SD = 16.55) and the 53.6% FSM cohort (M = 97.47, SD = 15.15), indicating that the lower FSM cohort performed significantly better on the test than the higher FSM cohort 9.6% cohort.

However, the 53.6% FSM group has a mean age of 6 months less than the 9.6% FSM group which could have affected this finding.

Cohort 2: Socio-economic status

The other groups that can be compared with an independent samples T test are:

Table 3.50 Cohort 2. The groups that are comparable for socio-economic status

North West (S)	14 (No. of children)	7-7:11	9.8%	Mean 106.71
North West (S)	10 (No. of children)	7-7:11	50.3%	Mean 98

Table 3.51 Cohort 2 group statistics

	FSM %age	N	Mean	Std. Deviation	Std. Error Mean
TotalTest	9.8	14	106.71	10.31	2.75
	50.3	10	98.00	10.98	3.47

Table 3.52 Cohort 2 Independent samples T test

		Levene's Test for Equal of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
TotalTest	Equal variances assumed	.227	.64	1.988	22	.06
	Equal variances not assumed			1.966	18.745	.064

Total Test Score analysis

The significance value for Levene’s Test for equality of variances is .64 (highlighted in yellow above). As this is larger than .05 equal variances can be assumed, so the first line of the table is used for the T test result.

The green highlighted T test result (.06) is more than .05, indicating that the difference in means between the two cohorts is not statistically significant.

There is no significant difference between the scores for the 9.8% FSM cohort (M = 106.7143, SD = 10.30630) and the 50.3% FSM cohort (M = 98, SD = 10.98484). However, the participant numbers in each of these cohorts are very small (14 and 10 respectively). This means that the power of the T Test is very low so this result is not a robust one.

The findings of the two post hoc tests for socio-economic status are, therefore, contradictory. These results will be discussed in more detail in the final discussion and conclusion chapter.

3.3.3.2 The effect of gender and age group on total test scores

A two-way between groups ANOVA was carried out to look at the effects of gender and age group (6-6:11; 7-7:11; 8-8:11; 9-9:11) on total test score (the dependent variable).

Table 3.53 Two-way ANOVA results showing the effect of gender and age group on total test score

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	26489.04[a]	7	3784.15	31.60	.00	.54
Intercept	2184050.00	1	2184050.00	18237.65	.00	.10
Gender	192.08	1	192.08	1.60	.21	.01
Year group	25696.16	3	8565.39	71.52	.00	.53
Gender * Year group	600.80	3	200.27	1.67	.17	.03
Error	22992.96	192	119.76			
Total	2233532.00	200				
Corrected Total	49482.00	199				

a. R Squared = .54 (Adjusted R Squared = .52)

Interaction effects

The Gender/Year group Significance is .17: $F = 1.67$ (both highlighted in yellow). As this is **more than** .05, this means there is no significant interaction effect.

The plot of the mean total test scores by the groups, age and gender can be seen below (figure 3.13). If an interaction effect was present (ie the lines in a graph of means were not parallel) the impact of one factor would depend on the other. For example, if male total scores declined with age, while for females it increased, then there would be an interaction effect.

The parallel or non-intersecting lines in the graph below indicate that there is no significant interaction between the two independent variables. This means that main effects can now be looked for.

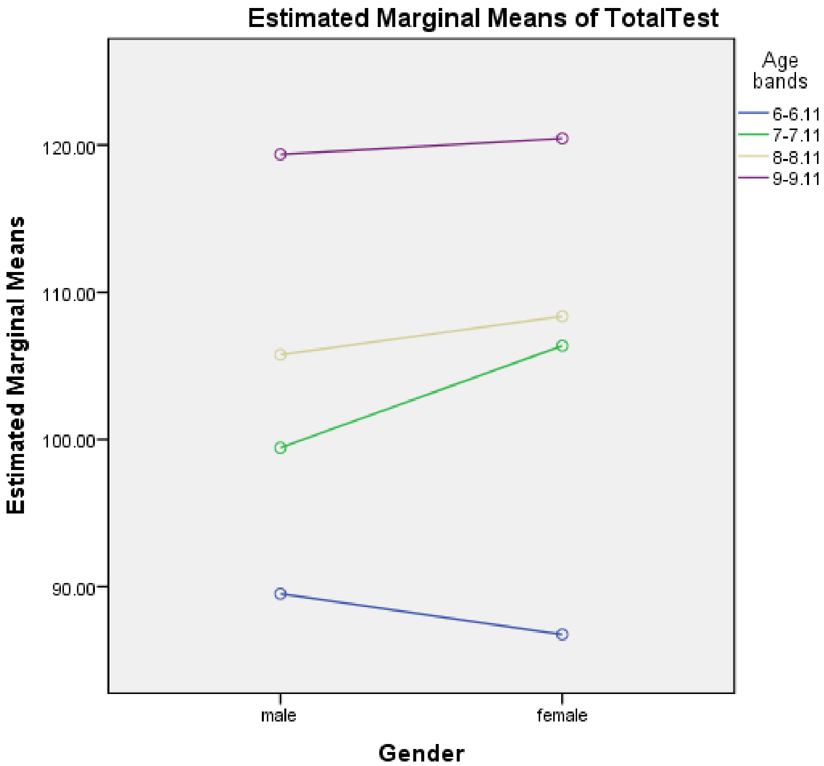


Figure 3.17 Marginal means diagram demonstrating no gender interaction effects

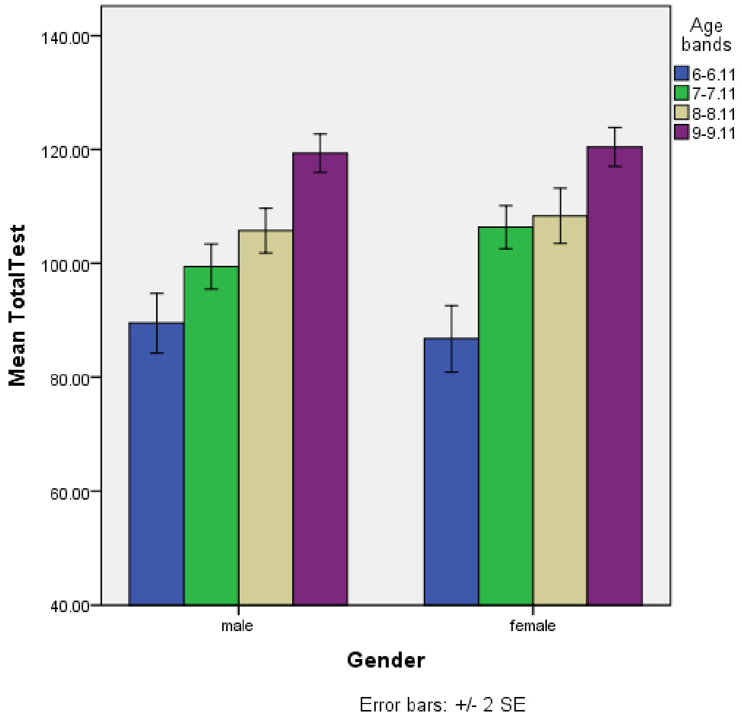


Figure 3.18 Bar chart summarising total test scores by age group and gender

Main effects for gender and year group

The green highlighting is explained below

Table 3.54 Levene's test of equality of variances

F	df1	df2	Sig.
2.884	7	192	.01

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Design: Intercept + Gender + Year grp + Gender * Year grp

As the significance for Levene's test of equality of error variances is .01 (highlighted in green above) and it needs to be more than .05 to mean that the variance of the dependent variable across the groups is not equal This means a more stringent significance level of greater than .01 needs to be set when evaluating the results of the 2 way ANOVA.

The significance figure for gender is .21: F=1.60) (both highlighted in green in table 3.53 above). This is more than .01, which means that no significant effect for gender was found.

The significance figure for year group is .00: F=71.52 (both highlighted in pink in table 3.53 above). This is not more than .01, so means that a significant effect for age group was found.
These results support the alternate hypothesis (Ha) in Hypothesis 1: that age affects test scores (ie the older age groups had higher total test score/ inferential and idiom comprehension than the younger age groups).
These results support the null hypothesis (Ho) in Hypothesis 2: that gender has no effect on the total test score.

3.4 Qualitative analysis of the test results

The main analysis of this test is quantitative. However, assessment of inferential comprehension is not a simple unitary process producing single correct responses. As explained in the methodology section, the majority of the test questions have a wide range of acceptable correct answers. This necessitated the drawing up of very detailed, comprehensive marking criteria (see appendix xi) to avoid as much subjective judgement as possible on the part of the marker. The current researcher opted for a correct/incorrect (1/0) scoring system. However, the child's response can be marked as incorrect for a variety of reasons, for example: giving a 'don't know' response, making a wrong inference, making a literal inference, and making an odd inference.
The data demonstrated that the younger children made many more literal interpretations than the older children (particularly in the idiom sub-section).
The typically-developing children in this study also gave some unexpected responses to some of the test questions. This is an important point to emphasize because unusual responses to inferential questions are usually attributed to children with communication impairments, in particular children with pragmatic difficulties. Some examples of these unusual responses are given below, with suggested explanations for the misinterpretations for those that had one. The age of the children is given in brackets after their answer (the full colour-coded list of immature or unusual responses is given in appendix xvii).

Immature or unusual responses to test questions

1) Simple Deductions

2. I am an animal You would see me in a zoo. I look like a horse. I have black and white stripes. What am I? (zebra)
Answer: A tiger (boy 6:09, girl 7:03, boy 8:11, girl 9:00). Presumably because of the stripes.
3. I am an animal I am pink or brown. I live on a farm in a sty. People say I am greedy. I have a curly tail. What am I? (pig)
Answer: A flamingo? Like a little bird (boy 6:01).

4. I am clothes. I am often made from wool. You put me on your hands. I keep your hands warm when it is cold. What am I? (gloves/ mittens/ muff)
Answer: A squirrel (boy 5:01).

5. I am food. I have a hard shell. You sometimes need to crack my shell to get at me. There are lots of different kinds of me. Squirrels like to eat me. What am I? (nut, any kind)
Answer: A lobster (girl 6:08). Presumably because it is hard-shelled food that you have to crack open.

6. You find me inside and outside buildings. I have a frame. You can see through me. If you throw a hard ball at me I could smash. What am I? (window)
Answer: A crocodile (boy 5:08).

8. I am made of metal and waterproof material I have a handle. You put me up and down. I keep you dry in the rain. What am I? (umbrella/ brolly)
Answer: A watering can (girl 6:01). She was distracted by another child running past.

2) Why Question inferences

3. Why do people have jobs?
Answer: So they get very clever (girl 5:07).
Answer: Because or they won't get to be parents (Boy 6:00).

4. Why mustn't you play with matches?
Answer: Cos you got to match them up again (boy 5:05). He understood a homonymn of match.

5. Why are mirrors made out of shiny surfaces not cardboard?
Answer: Because if it was too windy the cardboard would blow off it (boy 5:03).

7. Why are windows made out of glass not bricks?
Answer: Because they'll blow your house down (girl 5:03). Presumably influenced by the Three little Pigs story.

8. Why is a pan made out of metal not chocolate?
Answer: Because if you ate it you won't be able to draw on it (boy 5:07). He must have heard pan as pen.

11. Why are teeth covered in enamel not cotton wool?
Answer: Because they've got to move and you've got to go in the farm to walk properly (boy 5:01)

12. Why do we need pedestrian crossings on roads? (Pedestrian crossing described as 'like a zebra crossing' when she asked what it meant)
Answer: The cars stop and let the zebra past if it has escaped from the zoo (girl 6:09).

3) Making Predictions inferences
1. What would happen if you left a block of ice in the sun?
Answer: The world will be fire (boy 5:00)
Answer: You would be blind cos the sun would get in your eyes (boy 5:11). He may have heard ice as eyes.

5. What would happen if you took away the bottom tin in a stack of tins at the supermarket?
Answer: The pan would feel lonely (girl 5:05)

6. What would happen if you had a glass of water full up to the top and then you added 6 ice cubes?
Answer: It'll displode (girl 5:04)
Answer: You a cold brain (girl 5:10)

4) Formulating Solutions inferences (What could you do?)
2. What could you do if you wanted to buy a toy but you didn't have enough money?

Answer: Just count on from 10 to 11 (boy 5:10).
Answer: I would go and say to the witch people 'Please may I have some money' (girl, 6:08).

5. How could you get to the top floor of a very tall block of flats?
Answer: Wiv a ladder (boy 5:03).
Answer: You can go up in a helicopter and fly up (6:06).

7. What could you do if you burnt a cake you had baked for your mum's birthday?
Answer: Get the hose pipe and just put it out or if you had a fire extinguisher you could put it out (boy 6:04).

8. How do we keep milk fresh and cold?
Answer: Wiv the dryer (boy 5:03).

10. How could you talk to somebody who lives hundreds of miles away from you?
Answer: Drive to their house (girl 5:04).
Answer: You could copy their language – what they're saying (girl 5:05).
Answer: Go in an aeroplane (girl 5:10).

11. How could you find out what your teacher looked like when s/he was a little girl/boy?

Answer: Little (boy 7:04).
Answer: Contact the school (girl 9:02).

<u>5) Explaining inferences (How do you know?)</u>
1. How do you know that someone is tired?
Answer: Cos if their eyes are ugly (boy 5:07).
Answer: They're all lazy (boy 9:08).

4. How do you know that someone has been eating biscuits in the kitchen?
Answer: They'll be fat (boy 7:11)

6. How do you know that someone has put too much bubble bath in the bath?
Answer: It'll dust expode (just explode) (boy 5:00).

8. How do you know that someone has got toothache?
Answer: Cos they get all muddy (boy 5:01).
Answer: You'd have to get back in for your mum because your answer is 'No' (girl 5:03).
Answer: Because you can't hear them properly (girl 7:06). Presumably she is thinking of ear ache.

9. How do you know what's inside a box of something at the supermarket?
Answer: Use your senses. You could see a tail if it's a cat. You could like …. I don't know what else (girl 7:08).

10. How do you know that someone doesn't want you to hear what they are saying to someone else?
Answer: You can't have an attitood (attitude) (girl 5:05).

11. How do you know that someone thinks there is a bad smell in the room?
Answer: Someone had a flat water to add to a smelly room (boy 5:01).

<u>6) Making inferences from short passages</u>
2. Raj was playing on the beach. He trod on something and had to go to hospital? What did Raj tread on?
Answer: A hedgehog (boy 8:00).

3. Ellie paid the money to get in. She got changed and put her clothes in a locker. She put the key band on her wrist and took her towel with her. What was Ellie doing?
Answer: She was fixing the man (boy 5:05).
Answer: She was trying to get in (prompt) the ambulance (girl 5:10).
Answer: Maybe they had a pet and the pet escaped (girl 8:07).

4. Alisha and her mum went to see a concert but had to come back home without seeing it. What happened?
Answer: Her dog didn't come wiv her (girl 5:01).
Answer: Someone come and wobbed the cookies and all the food so they can't get no cash (girl 5:03).

5. Dan's dad took him to somewhere outside. He pushed Dan up and down on something and pushed him round and round on something else. Dan climbed up some steps and went down something. Where were Dan and his dad?
Answer: In the doctor's (boy 5:05).
Answer: The attic (boy 5:08).
Answer: On the lift (girl 5:10).
Answer: In the basement (8:00).

6. A group of friends went out to a party together. They had planned to get the bus home afterwards but ended up taking a taxi. Why?
Answer: Buses are tall, taxis are small (boy 8:05).

7. Jane and her friend Sasha went out for a bike ride. They came back later than they were supposed to and Jane was pushing her bike. Why?
Answer: Cos she was sleeping (boy 5:08).

8. James and his dad are sitting in the front row. The curtain goes up and people in costumes come on and talk loudly and sing. Where are James and his dad?
Answer: They're in the beach and the park and they fell off and Izzy came over to say 'What've you been doing?'(girl 5:03).
Answer: Cos they were on Halloween (boy 5:10).
Answer: On the bus (girl 6:02).

9. Mum and dad have just eaten a nice meal in an expensive restaurant. They ask for the bill and then both look worried and embarrassed when it comes. Why?
Answer: Because they're so young (girl 5:04).
Answer: In case grandma and grandad are looking after the children fought that they could have afters. Actually, I don't know (girl 5:05).
Answer: Cos they might have done something the police wanted them to do (boy 5:08).

10. Dad bought a chicken to cook for Sunday dinner but the family ended up eating just vegetables and gravy. What happened?
Answer: Chicken ran away (boy 5:01).
Answer: They were all chubby (boy 5:05).
Answer: It could be Monday (girl 7:08).

<u>7) Situation-based emotions inferences</u>
9. A boy hits Gemma's sister at school. How does Gemma feel? (angry).
Answer: Happy if she hates her sister or angry (boy 9:06).
10. Jamie's dad takes him for tea at McDonalds but McDonalds is closed. How does Jamie feel? (sad).
Answer: Scared (girl 5:10).
<u>8) Belief-based emotions inferences</u>
Samia wants to go clothes shopping with her mum and she will be going clothes shopping.
Samia thinks she is going to help her mum clean the house.
7) How does she feel? (sad).

Answer: Happy that she's going, but not happy that she thinks she will be helping her mum do the washing – was it washing? (boy 7:10).

10) Strange Stories

Pretend

John and Rashid are playing spacemen. John picks up a bowl and puts it on his head. He says: 'This bowl is a space helmet.' 1) Is it true what John says? 2) Why does he say it?

Answer: Because he's imaginating (girl, 7:02).

Answer: Because they don't wear that otherwise they'd die – no oxygen (boy 8:11).

Lying

Julie hates going to the doctor. She has very bad stomach pains and is holding her stomach. Her mum asks 'Have you got tummy ache?' Julie says 'No'. 3) Is it true what Julie says? 4) Why does she say it?

Answer: So she can carry on going to the shops (boy 8:05).

White lie

Tom's friend Samia has just had her hair cut. Tom doesn't like it. He doesn't think it suits her. He liked her old style better. When Samia asks Tom 'Do you like my new hair style? Tom says 'Yes, it's very nice.'11) Is it true what Tom said? 12) Why did he say it?

Answer: Tos (cos) if he said 'No' they would like punch him (boy 6:06).

Answer: So then she'll think he likes the gift (boy 8:11).

11) Faux pas

Vase

Mrs Patel has Mrs Brown, an old friend, over for tea. She hasn't seen Mrs Brown for years. Mrs Brown accidentally knocks over a vase and breaks it. Mrs Patel says 'It doesn't matter. It was a wedding present that I never liked anyway. I can't remember who bought it for us." Mrs Brown says 'I bought it for you.' 10) How does Mrs Brown feel? 11) Did Mrs Patel remember that Mrs Brown had bought her the vase? 12) How does Mrs Patel feel?

Answer: She might be just eatin her dinner or laughing (Boy 6:00).

12) Idioms

1. Keep your eyes peeled.

Answer: It means that there's rats or monsters around (boy 9:05).

2.I've got butterflies in my tummy.

Answer: When you eat lots of sweets (boy 6:09).

3.My lips are sealed.

Answer: Like they're chapped and stuck together (boy 8:04).

4. Let's hit the road.

Answer: Let's det (get) the party started (boy 6:06).

5. Put a sock in it.

Answer: It'll be naughty. That's why. What happened. Somebody dies and it goes boom and the car – it's fell over (Boy 5:00)
Answer: You have to dig it up (boy 6:11).

7. She's over the moon.
Answer: A cow jumped over the moon (girl 6:06).
Answer: She's on fire – answers all the questions right (girl 7:06).
Answer: She's really clever (boy 7:08).
Answer: She's really far away (girl 9:06).

8. He's having a whale of a time.
Answer: She wants to go away for a little bit (girl 6:02).

9. She got out of the wrong side of bed today.
Answer: Maybe she was too early for school and no one was there (boy 6-6:11).
Answer: She's a lunatic (boy 9:06).

12. It's not my cup of tea.
Answer: It's not your teabag (boy 7:05).
Answer: I don't really like it – I prefer black coffee (girl 9:06).

14. He's going to turn over a new leaf.
Answer: He's going to have a baby. He might be married (girl 7:03).

15. You're pulling my leg.
Answer: You're holding me back (girl 9:09).

16. I'm tied up at the moment.
Answer: Someone ties them up in prison after getting in a matrix van with other people (girl 6:05).

17. She'll hit the roof.
Answer: You can walk in it when you're doing house work (boy 6:11).
Answer: Go really fast at work (girl 7-7:11).
Answer: She's big (girl 8:09).
Answer: She's jumping on the trampoline in the house (girl 9:05).

19. It goes in one ear and out the other.
Answer: Earring (girl 6:02, girl 7:09).
Answer: A piece of string (boy 6:06).
Answer: A machine (girl 6:06).
Answer: A phone (boy 6:09).
Answer: She's magic (boy 7:01) Magic (girl 7:07).
Answer: It means it goes through that one and out that one, like ear wax (boy 7:09).
Answer: One kind is funny and one kind isn't. That was just a guess (boy 7:10).
Answer: I'm not really sure. It's something goes round people (girl 8:02).

20. You've hit the nail on the head.
Answer: She's saying you're grounded (boy 7:01).

Answer: You've made me shocked (girl 7:06).
Answer: You've give me a headache (boy 7:09).

3.5 <u>Four exploratory case studies</u>

<u>Using the HICIT to assess children with communication impairments</u>

Four children with communication impairments were assessed with the final study version of the HICIT by the current researcher. These were: A. a 5:06 year-old boy with high functioning ASD (HFA), B. a 6:03 year-old boy with HFA, C. a 10:02 year-old boy with DiGeorge syndrome, and D. a 9:10 year-old boy with HFA. Child A and child D, both with HFA, were assessed as part of the test standardisation process in one school and child C as part of the test standardisation process in another school. Their data were not included in the collation of the scores as their profiles fell within the exclusion criteria. Child B and C attended resourced units in two different mainstream schools in the North West which were also included in the standardisation procedure. These boys were assessed at a later date however. Their parents were informed verbally about the assessment and were asked to give their written consent for the child to be assessed and video-recorded (for teaching and research purposes) by completing and returning the form in appendix xix.

Three of the four boys scored well below the mean score for their chronological aged peers on the HICIT and the fourth boy scored well below his comprehension of grammar age. They also produced some very unusual qualitative responses to many of the test questions. Selected examples of these are given below but more are given in appendix xviii.

<u>Child A</u>

Aged 5:06, child A demonstrated good comprehension of grammatical structures, obtaining an Age Equivalent of 6:06 (Standard Score 113, 81st percentile) on the Test for Reception of Grammar 2 (TROG-2). He also scored age appropriately on the information and grammar sections of the Renfrew Action Picture Test, an expressive language test.

On the HICIT he scored **26/84**. This is well below the mean of **48.12** for 5:00-5:11 year-olds and well below the lowest score in the range for boys (32-64). He had severe difficulties in all of the sub-tests apart from the first one (making deductions, where he got 5/8). In view of his more than age appropriate comprehension of grammatical structures on the TROG-2 this HICIT test result indicates that he has a specific problem with inferential comprehension.

Some of his very unusual responses are given in blue below with responses from typically developing 5:06 year-olds provided alongside them.

1) Simple deductions

1. I am a musical instrument. You bang me with your hand or with sticks. I make loud noise. What am I?

Child with HFA (A): A lady. Typical Child (TC): A drum.

2) Why Question inferences

4. Why mustn't you play with matches? TC: Because they can burn you. Because you've got to match them up (boy 5:05). A: Because they have to match together.

12. Why do we need pedestrian crossings on roads? (Explained like a zebra crossing) TC: Don't know; so you don't go off the road. A: The zebra crossings are roads that are for zebras to cross – to lead the zebras back to the zoo.

3) Making predictions inferences

4. Sam and Ruby had a date to go to the cinema but Ruby didn't turn up How did Sam feel?

TC: Upset/sad. A: Great.

6. What would happen if you had a glass of water full up to the top and then you added 6 ice cubes?

TC: It'll spill over; it'll melt; don't know. A: Because the 6 water puddles have to get water to buy them 1,2,3,4 and 5 series Moshlings.

4) Formulating solutions inferences (What would you do?)

11. How could you find out what your teacher looked like when she was a little girl? TC: Don't know. A: She looked like a bumper girl.

5) Explaining inferences (How do you know?)

2. How do you know that a group of children want you to come and join in their game?

TC: Don't know; they ask you. A: Because I have to play with them.

A's answers provide rich qualitative information. He interprets the first question literally, telling me I am a lady rather than realising he has to deduce what the object is from the given clues. He gives many answers related to his special interests (numbers and Moshlings). He demonstrates difficulty with empathy and social interaction enjoyment. These are all recognised characteristics of children with ASD. The typically-developing 5:06 year-olds use the 'don't know' response much more than A does.

<u>Child B</u>

Child B with HFA attends a resource base for children with ASD half-time and a mainstream year 1 class half time. He was 6:03 when assessed. He scored an Age Equivalent of 5:03 (Standard score 88, 21st centile) on the TROG-2 indicating that comprehension of grammatical structures was delayed but within normal limits (WNL) for his age (Owens, 2004). He scored a total of **31** on the HICIT. This is well below the average score for 6:00-6:11 year-olds of **88.** This indicates that he has a specific difficulty with inferential comprehension. He had severe difficulty understanding all of the sub-sections apart from section 7 (situation-based emotions) where he scored 7/16. Some examples of his unusual responses are given in blue below:

<u>1) Simple Deductions</u>

6. You find me inside and outside buildings. I have a frame. You can see through me. If you throw a hard ball at me I could smash. What am I? B: Naughty

<u>2) Why Question Inferences</u>

10. Why shouldn't you agree to take a lift in a car from people you don't know? B: People can't have triangle wheels or rectangle wheels and you won't drive it. Or squares. You need circly wheels to drive it so it can move properly.

<u>3) Making Predictions Inferences</u>

1. What would happen if you left a block of ice in the sun? B: The sun will be angry and then it will shine someone.

<u>4. Formulating Solutions Inferences (What could you do?)</u>

5. How could you get to the top floor of a very tall block of flats? B: But you can't get everything balanced on top. You can't go to floor 10 because there's no floor 10 in this world.

<u>5. Explaining Inferences (How do you know?)</u>

2. How do you know that a group of children want you to come and join in their game? B: But any people doesn't want any games. But the boy doesn't want to play the game but the rest of the people want to.

6) Making Inferences from short passages

3. Ellie paid the money to get in. She got changed and put her clothes in a locker. She put the key band on her wrist and took her towel with her.

What was Ellie doing? B: Ellie is a good girl but he can't walk hisself but he is 4 years-old. When you are a 4 year girl you can't go by yourself. You need an adult.

7) Situation-based emotions inferences

8. Three year-old Tom has lost his mum in a busy shopping centre. How does he feel? (scared). B: Happy

14. Jake has won first prize in a competition. How does he feel? (happy). B: Scared

10) Strange Stories - White lie

Tom's friend Samia has just had her hair cut. Tom doesn't like it. He doesn't think it suits her. He liked her old style better. When Samia asks Tom 'Do you like my new hair style? Tom says 'Yes, it's very nice.' 11) Is it true what Tom said? B: Yes.

12) Why did he say it? B: Because the other boy doesn't like hair cutting. He likes to grow it back all day.

11) Faux pas - Disco

It was the school disco. Jill was in one of the toilets. Karen and Sara came in to the toilets afterwards. Karen said: 'Did you see how Jill was dancing. Wasn't she terrible!' Then Jill came out of the toilet. 1) How does Karen feel? B: Sad because she missed the thing out of the dance. 2) How does Jill feel? B: So happy. 3) Did Karen know that Jill was in one of the toilets? B: He was human, but he had to go to the toilet very quick.

12) Idioms

4. Let's hit the road. B: That means everything will be damaged and you can't go back in earth again, but you can make a new earth and astronauts can have for them.

7. She's over the moon. B: She is over the moon but that's because it is all icy and there is cheese in it.

These qualitative responses indicate some special interests (shapes and numbers), some difficulties interpreting basic emotions, empathy and social understanding difficulties and pronoun confusion (she becoming he). These are typical features of children with ASD.

Child C

Child C has DiGeorge Syndrome. This is a congenital disorder caused by the deletion of chromosome 22q11. Features of the condition include a learning disability and pragmatic language impairment. C attends a resource base for children with moderate learning disabilities attached to a mainstream school. He was 10:02 when assessed. He obtained an Age Equivalent of 6 years (5th Centile) on the TROG 1. This indicates significantly delayed comprehension of grammatical structures for his chronological age. However, C has a moderate learning disability as part of his syndrome so the score is in line with his cognitive age. He obtained a total test score of **50** on the HICIT. This is well below the mean score for 6:00-6:11 year-olds (**88**). These results indicate that C has got delayed understanding of grammatical structures (in line with his developmental age) but he has even more of a delay in his inferential comprehension.

He scored well below the mean for 6 year-olds in all sections apart from deductions; formulating solutions and situation-based emotions. Some examples of his more unusual responses are given in blue below:

2) Why Question Inferences

9. Why are wellies made out of rubber or plastic not paper? C: Cos they need to change all my underwear. Because if you um done an accident you need to get change and tell your mummy and daddy or auntie and uncle.

3) Making Predictions Inferences

5. What would happen if you took away the bottom tin in a stack of tins at the supermarket? C: Er – they um you might get shouted ….. then the police might come and take you to jail.

5. Explaining Inferences (How do you know?)

4. How do you know that someone has been eating biscuits in the kitchen? C: Er – you buy more

6. How do you know that someone has put too much bubble bath in the bath? C: You take the plug out

6) Making Inferences from short passages

7. Jane and her friend Sasha went out for a bike ride. They came back later than they were supposed to and Jane was pushing her bike. Why? C: Um – (repeat). Cos um

maybe because her pet was in the – you know the way …. Is it a girl or a boy? … her pet was in the way…. My girlfriend has a pet. She has a little puppy and her name is Poppy.

10) Strange Stories - White lie (scenario above)

11) Is it true what Tom said? C: Yep

12) Why did he say it? C: Cos it was beautiful. Yep yep yep

11) Faux pas - Disco (scenario above)

1) How does Karen feel? C: Oh you mean like Karen from sponge Bob? Um happy

2) How does Jill feel? C: Um – kinda like dancing like toy Barney. (Rpt) Happy

3) Did Karen know that Jill was in one of the toilets? C: Yeh.

12) Idioms

2. I've got butterflies in my tummy. C: Er. It means she might be sick like this (demonstrates). I be like that all week.

4. Let's hit the road. C: If you try to hit the road then if you step on it all the time than a car or a bus or a taxi might run over you and you have to go home and have a plaster on it.

10. He's driving me round the bend. C: Cos he might be trying to take theirselves to the cinema to see the new Sponge Bob movie.

These qualitative responses indicate some special interests (Sponge Bob), pronoun confusion and literal interpretation of idioms. These are typical features of children with ASD. He also had particular difficulty answering the 'How do you know? questions which involve theory of mind. He answered these as if they were 'what could you do? questions. This could be because the previous section was that type of question. However, he was given some models and extensive prompts with this section and he was still only able to answer 1/12 correctly.

Child D

Child D with HFA attended a mainstream school. He was 9:10 when assessed. He passed 19 complete blocks on the TROG-2, giving him an Age Equivalent score of over 12 years (standard score 111, 77th centile). His total test score on the HICIT was 104. The mean for the 9:00 to 9:11 age group is 120. He scored at the lowest end of the range for this age group. However, his comprehension of grammar is significantly more advanced than his inferential comprehension (by 3 years). His main difficulties on the HICIT were with the belief-based emotions, faux pas and idioms.

3.6 Chapter summary

This chapter has outlined the pilot study results and how the pilot assessment was converted to the final assessment. Tests of validity (exploratory and confirmatory factor analysis) and reliability (Cronbach's Alpha, Rasch analysis and inter-rater reliability) were reported. These demonstrated a single factor loading and good model fit measurements. The reliability of the test, as assessed by Cronbach's Alpha was moderate. Rasch analysis indicated that the internal consistency of the test was good. The inter-rater reliability of the HICIT scoring was 98.6%.
A summary of the final study results was provided. This was sub-divided into: participant information, descriptive statistics and inferential statistics. The inferential statistics tested the hypotheses that age, socio-economic factors and gender affected the HICIT scores. The results indicated that there was a developmental progression between the age groups for inferential and idiom comprehension but that there was no effect for gender. The idioms sub-section was the only section not to reach ceiling scores by 9:11. The results for the relationship between socio-economic status and test scores were contradictory and therefore inconclusive. Qualitative analysis of the results was detailed in the penultimate section. This illustrated developmental differences in the types of responses given by typically-developing children. The final section summarised the results of case studies of the HICIT being used to assess four children with communication impairments.

The next discussion and conclusion chapter (chapter 4) will discuss these results in more detail, draw conclusions from them and compare them to the results for other existing standardised language assessments.

CHAPTER FOUR: DISCUSSION AND CONCLUSIONS

Overview of the chapter

This concluding chapter revisits the study aims and hypotheses and discusses whether these have been met. An examination of the test's validity and reliability is carried out and the results compared to validity and reliability results from other existing language assessments. Normative data summaries of the final HICIT assessment are included covering the effects of age, socio-economic status and gender. The chapter also includes a discussion of the results of the exploratory case studies on the use of the HICIT assessment with four children with communication impairments. Qualitative analysis of the study results is discussed in detail. A summary of the limitations and advantages of the HICIT is given. Suggested amendments to the test are given and areas for future research recommended. Feedback on the HICIT from practising clinicians is outlined. Finally, the steps needed to get the test published are detailed.

A reminder of the study's primary and secondary aims and hypotheses is given below:

Primary Study Aims and related Hypothesis:

1) To develop a robust assessment of inferential and idiom comprehension for 5:00-9:11 year-old children and to provide validity and reliability data supporting the developed assessment.

2) To provide normative and statistically significant data for inferential and idiom comprehension in typically developing children aged 5:00 to 9:11.

   Related Hypothesis

   Null Hypothesis (Ho): The age group of the participant will have no effect on inferential and idiom comprehension scores.

   Alternate Hypothesis (H1): The older age groups will have higher inferential and idiom comprehension scores.

3) To qualitatively analyse the developmental progression of inferential and idiom comprehension.

4) To carry out exploratory case studies using the HICIT with children with communication impairments to determine if it could be a useful assessment tool to assist in differential diagnosis.

Secondary Study Hypotheses

1)To compare inferential and idiom comprehension between genders; to determine if there is a statistically significant difference between female and male participants.

Null Hypothesis (Ho): The gender (female or male) of the participant will have no effect on inferential and idiom comprehension scores.

Alternate Hypothesis (H1): The gender (female or male) will have an effect on inferential and idiom comprehension scores.

2) To compare inferential and idiom comprehension to the socio-economic status of the participants to see if there is a statistically significant difference between children from different socio-economic backgrounds.

Null Hypotheses (Ho): The socio-economic status of the participants will have no effect on inferential and idiom comprehension scores.

Alternate Hypothesis (H1): The socio-economic status of the participant will have an effect on inferential and idiom comprehension scores.

Bishop (1997) emphasises that developing a standardised language test is an extremely time-consuming and expensive business. She reports that many frequently used language assessments are poorly standardised and have poor reliability. Brandel et al (2011) state that there is limited research evidence on the comparison of different language tests and the interpretation of their results to aid differential diagnosis. They highlight the fact that many pre-school language assessments have been compared but very few school-age language tests, even though these are regularly used to support clinical decisions about appropriate intervention. They found inconsistencies in the reliability information and the scores of children on two school age language assessment that they compared (the TOLD-P:3, Newcomer and Hammill 1997, and the CASL, Carrow-Woolfolk, 1999). Their findings emphasized the need to select assessments based on the theoretical constructs that best fit with the client group being assessed. They also encouraged the use of multiple measures to assess language skills. The HICIT could become one of these multiple measures to be used when assessing children with high level comprehension difficulties.

All of the study aims set at the beginning of the HICIT project were met. They are discussed in turn below:

4.1 Primary study aim 1

To develop a robust assessment of inferential and idiom comprehension for 5:00-9:11 year-old children and to provide validity and reliability data supporting the developed assessment.

The study's validity results will be discussed first and then the reliability results:

4.1.1 Validity of the HICIT

Three main types of validity will be discussed: construct validity, content validity and concurrent validity.

4.1.1.1 Construct validity

One method of examining a test's construct validity is by measuring the internal consistency of the sub-tests. This measures the degree to which the individual test items are measuring the same underlying dimension (Adams et al, 2001). As this is measured by using a reliability coefficient (Cronbach's Alpha), this aspect of construct validity will be considered in the following reliability section.

The main other statistical methods used to examine construct validity of the HICIT were exploratory and confirmatory factor analysis. McCauley (2001) states that when a target measure shares (or 'loads on') the same factor as has been demonstrated to be valid elsewhere, then construct validity is demonstrated.

Exploratory Factor Analysis (EFA)

EFA was carried out on the full data set sample of 200 children, aged 6:00 to 9:11 to determine the number and nature of common hypothetical underlying (or latent) factors that underpin the observed variables (Hatcher, 2013). Only the first component had an Eigenvalue of 1 or more (5.56), indicating that all of the items in the HICIT load on to one underlying factor. This is a positive finding, as all of the sub-tests are based on the developmental literature on inferential and idiom comprehension so you would expect them to have the same underlying construct. The sub-tests do not aim to assess any other aspects of verbal language such as vocabulary or syntax.

Unrotated loadings of the sub-tests (the component matrix values) on this first factor were all above .4, which is quite a strong loading (Pallant, 2013). However, the weakest loadings were on the belief based emotions sub-test (.49) and the mental state verbs (.55).

Confirmatory Factor Analysis (CFA)

CFA was carried out on the full data set sample of 200 children, aged 6:00 to 9:11 to determine if the constructed HICIT model is a ‘good fit’ or not and to confirm or refute specific hypotheses concerning the structure underlying a set of variables (Pallant, 2013). The analysis was carried out on all 12 sub-tests, then with the sub-test with the weakest standardised estimate (SEM) loading removed (belief based emotions, SEM .44), and then with belief-based emotions still removed and the sub-tests which had the largest modification index (MI) (between e7 and e12, MI 10.22) co-varied. These modifications only slightly improved the SEM values in the remaining sub-tests. Table 3.8 in the results section indicates that the SEM measures from the CFA and the component matrix values from the EFA were very similar, the EFA values being just .02 to .05 more than the CFA values. Table 3.8 demonstrates that the sub-sections with the highest loadings (over .7) are idioms (CFA .77, EFA .79), inferences from passages (CFA .77, EFA .79), explaining inferences (CFA .74, EFA .76) and strange stories (CFA .71, EFA .73). Those with the lowest loadings (.5 or below) are belief-based emotions (CFA .44, EFA .49) and mental state verbs (CFA .5, EFA .55).

These results give further support to the removal of section 8 (belief-based emotions) from the final version of the HICIT.

The model fit measurements CMIN/DF, GFI, AGFI, CFI and RMSEA (see 3.12 in the results chapter) all indicate that the model is a good fit apart from the probability (P = .14) value for the third model (with the belief-based emotions sub-test removed and the e7 and e12 co-varied). The best P figure (P = .02) is for the second model, with the belief-based emotions sub-test removed. This provides further support for the removal of sub-test 8 from the final version of the HICIT.

Brandel et al (2011) define construct validity as a test's ability to accurately reflect the conceptual foundation upon which it has been developed. That is, if a test has good construct validity, it appears to be assessing the abilities that it was designed to measure. They stress the importance of including correct information about construct validity in the test manual. In their randomised control study of language impaired children they carried out unrotated principle-component factor analysis on the CASL (Carrow-Woolfolk, 1999) and the TOLD-P:3 (Newcomer and Hammill, 1997). The CASL manual reported a 3 factor loading model of language ability (form, content and use). However, Brandel et al (2011) found only a single factor loading with this test. They used this as evidence for the theoretical construct of the unidimensionality of language during the elementary years.

The TOLD –P:3 manual reported that the test loaded on a single factor but Brandel et al (2011) found that it loaded on 2 separate factors. The sub-tests that loaded onto the first factor required minimal responses from the children (pointing to a picture, filling in the last word of a sentence, or imitating an utterance). The sub-tests that loaded onto the second factor required more complex responses and the construction and coordination of multiple utterances. They suggested the first factor could be named ‘Utterance Level Communication’ whereas the second could be named ‘Discourse Level Communication’. They used this as evidence that underlying theoretical language constructs could be more aligned with language-processing load than with the traditional constructs of receptive versus expressive language skills or linguistic subsystems including semantics, morphology, and syntax. The results of factor analyses on both language assessments did not support the comprehension versus production dichotomy of language abilities. The authors encouraged additional research to resolve these language construct and measurement issues.

The single factor measurement of the HICIT supports the theory that inferential and idiom comprehension are a unified underlying theoretical linguistic concept.

4.1.1.2 Content validity

Content validity or logical validity refers to the extent to which a measure represents all components of a given construct. It requires the use of recognized subject matter experts to evaluate whether test items assess defined content. In clinical settings, content validity refers to the correspondence between test items and the features of a disorder. There is a statistical method for assessing content validity by getting subject matter experts to rate how essential the items of tests are and then applying a formula to those results (Lawshe, 1975). This statistical analysis was not used for the HICIT due to time constraints. Instead the current researcher followed Adams et al’s (2001) content validity specifications from when they developed the ACE. That is, the tasks and content of the HICIT were guided by the interests and language abilities of 5 to 11 year-old typically-developing children, culture-specific items were not included in the assessment, and speech and language therapy colleagues gave advice on the suitability of the items in the initial stages of test construction.

4.1.1.3 Concurrent validity

Concurrent validity refers to the consistency of results between two tests that assess the same skill. It is assessed by comparing the standard scores of participants on similar assessments. Brandel et al (2011) point out that this involves extensive testing and data collection for a large number of children so it is very expensive and time consuming. Consequently, concurrent validity is not often reported upon in the language assessment development literature. Adams et al (2001) did measure concurrent validity after they had developed the Assessment of Comprehension and Expression (ACE) assessment. They assessed 163 children on the ACE (Adams et al, 2001), the Test for Reception of Grammar (TROG, Bishop, 1989) and the British Picture Vocabulary Scales, Second Edition (Dunn et al, 1998.). Individual children were assessed with all three tests during the same week. They found that the mean standard scores obtained on all three assessments were very similar, demonstrating very good agreement between them. Brandel et al (2011) conversely did not get such a close correlation of marks when they assessed 216 American children with the TOLD-P:3 (Newcomer and Hammill, 1997) and the CASL (Carrow-Woolfolk, 1999) within a three month period. They used the Pearson r coefficient and a paired samples T-test to compare the children's performance on both tests. Analysis of the results revealed that the concurrent validity pattern differed from those found in the norming samples as cited in examiner manuals. They found that only 64% of the children classified as having language impairment on the basis of their performance on the TOLD-P:3 were also classified as having a language impairment on the basis of their performance on the CASL.

The current researcher plans to assess the con-current validity of the HICIT once the test has been finalised. It will need to be compared to other standardised British assessments. However, it would not be appropriate to compare it to the TROG, a test of understanding of grammatical structures, as one of the reasons for developing the HICIT was because a cohort of language-impaired children scored within normal limits on the TROG but still displayed significant difficulties with inferential and idiom comprehension. The BPVS is a possible receptive assessment for comparison, but from the literature review results presented in chapter one, receptive vocabulary would not necessarily be expected to correlate well with inferential and idiom comprehension. If it did correlate well with inferential comprehension it would fit in with Lucas and Norbury's (2015) finding that it is vocabulary that matters most in inferential comprehension. The most appropriate measures to compare the HICIT to would be the inferential comprehension and non-literal comprehension sub-sections of the ACE (Adams et al, 2001) as these are the closest to the type of verbal skills assessed in the HICIT. If a correlation were found between the HICIT and these ACE sub-tests it would support Norbury and Bishop's (2002) theory that inferential comprehension is an independent sub-type of comprehension.

4.1.2 Reliability of the HICIT

The three main types of reliability: internal consistency, inter-rater reliability and test-retest reliability (McCauley, 2001) are considered in turn below:

4.1.2.1 Internal consistency reliability: Cronbach's coefficient Alpha

The most commonly used statistical measure of internal consistency reliability is Cronbach's coefficient Alpha. This was briefly mentioned in the construct validity section above but will be considered in detail here. The higher the value of Alpha (which goes from 0-1), the more reliable the items. McCauley (2001) defines a value of less than .2 as a slight, negligible correlation; .2-.4 as a low correlation; .4-.7 as a moderate correlation; .7-.9 as a high correlation; and over .9 as a very high correlation. However, she also stresses that a strong correlation between factors is not evidence of a causal relationship between them (McCauley, 2001).

The reliability of the sub-test items across the age groups will be considered in turn:

Deductions – Cronbach's Alpha (CA) varies from .15 (negligible) in the 7:00-7:11 age group up to .49 (moderate) in the 5:00-5:11 age group. This could be evidence to use this sub-test for the younger children only. However, this sub-test has only 8 items. Briggs and Check (1986, in Pallant, 2013) state that the CA value can be quite small for scales of less than 10, so this may explain the very low correlation scores. In addition, it is useful to keep a relatively easy sub-section at the beginning of the test as it is a good warm up exercise and confidence booster for the children of all ages. Consequently, it will be maintained it its current form.

Why Questions – CA varies from .18 (negligible) in the 8:00-8:11 age group up to .56 in the 5:00-5:11 age group Although these are not strong correlation scores the mean scores in this sub-test increase for each age group (from 6.6 at 5:00-5:11 to 11 by 9:00-9:11). Consequently this sub-test will be kept in its entirety.

Predictions – has the lowest CA scores of all the sub-sections ranging from -.66 in the 8:00-8:11 age group to .11 (negligible) in the 5:00-5:11 age group

This sub-test has only 6 items which Briggs and Check (1986, in Pallant 2013) would say could explain the very low correlation scores. In addition, this sub-test was created by amalgamating two different sub-tests from the pilot version of the HICIT (making predictions and taking others' perspectives). The poor reliability score could indicate that different constructs are being assessed. This result adds evidence to the suggestion that this sub-section should only be used for the two younger age groups (5:00-6:11).

Formulating solutions – CA varies from .30 (low) in the 7:00-7:11 age group to .7 (high) in the 6:00-6:11 age group As the latter age group has the only high correlation and the 5:00-5:11 age group the next highest (.61), this result adds evidence to the suggestion that this sub-section should only be used for the two younger age groups (5:00-6:11).

Making inferences – CA varies from .41 (moderate) in the 9:00-9:11 age group to .74 (high) in the 5:00-5:11 age group In addition, the mean scores increase from 5.1 in the youngest children to 11.2 in the oldest. This sub-test will be maintained in its entirety.

Situation-based emotions – CA varies from .4 (low-moderate) in the 8:00-8:11 age group to .54 (moderate) in the 5:00-5:11 age group The mean scores rise from 9.6 in the youngest to 14.8 in the oldest so this sub-test will be maintained in its entirety.

Belief-based emotions - CA varies from .27 (low) in the 6:00-6:11 age group to .51 (moderate) in the 8:00-8:11 age group This sub-test has only 8 items which Briggs and Check (1986, in Pallant 2013) would say could explain the low correlation scores. However, as the mean scores increase by less than two marks from youngest to oldest age group and due to the 50% chance of guessing the correct answer, this sub-section will be removed from the final HICIT.

Mental state verbs - CA varies from .45 (moderate) in the 9:00-9:11 age group to .67 (moderate) in the 7:00-7:11 age group These moderate reliability scores justify this sub-section being kept in. However, as the mean scores for the 7:00-7:11 and the 8:00-8:11 year groups are nearly the same, these two age groups could be joined together in the final HICIT.

Strange Stories - CA varies from .30 (low) in the 8:00-8:11 age group to .67 (moderate) in the 7:00-7:11age group As the mean score for the 7:00-7:11 age group is slightly higher than the mean score for the 8:00-8:11 year-olds, if these two age groups are amalgamated in the final HICIT, this could improve the CA score for that joint age group

Faux pas - CA varies from .52 (moderate) in the 8:00-8:11 age group to .67 (moderate) in the 6:00-6:11 age group These results justify the sub-test being kept in the final HICIT. The mean scores for the 7:00-7:11 and the 8:00-8:11 year groups are nearly the same, so these two age groups could be amalgamated in the final HICIT.

Idioms – This sub-test has the highest overall reliability scores. CA varies from .55 (moderate) in the 6:00-6:11 age group to .73 (high) in the 8:00-8:11 age group The mean scores rise each year from 3.2 in the 6:00-6:11 age group to 11.4 in the 9:00-9:11 year-group The top mean score is only just over half of the possible maximum score of 20, indicating that this area of comprehension is still developing at 9:11. This justified the idiom extension study detailed in section 4.21 below.

The CA averages across the ages for the sub-tests in order of highest to lowest reliability are: Idioms .65; Faux Pas .59; Making inferences .57; Mental state verbs .56; Strange stories .54; Formulating solutions .53; Making inferences from short passages .46; Situation-based emotions .46; Deductions .37; Belief-based emotions .37; Why questions .35; Predictions .13. This ranges from moderate to negligible reliability according to McCauley (2001).

This is a slightly different order for the exploratory and confirmatory factor listings but idioms and making inferences are highly ranked and belief-based emotions lowly ranked in both systems.

Comparative findings from other language tests

Other standardised language assessments report better CA scores than the HICIT CA scores. For example the Test for Language Competence (Wiig and Secord, 1989) gives a reliability score of .86-.92 for level 1 (age 5-9:11) and .75-.82 for level 2 (age 9-18:11). The CASL (Carrow-Woolfolk, 1999) reports moderate to strong reliability scores of .64 to .94. However, these are figures based on the whole assessment which covers many components of language: semantics, syntax and comprehension. Many sub-components of language are also assessed such as concepts, grammatical structures and more abstract language (including inferences and idioms). Consequently it is not equitable to compare these CA figures to the HICIT CA figures. A more appropriate comparison would be to the sub-tests of the British language assessment, the ACE (Adams et al, 2001), that is to the inferential comprehension sub-test and the non-literal comprehension sub-test.

The making inferences sub-test of the HICIT can be compared to the inferential comprehension sub-test of the ACE. The average CA scores across the ages for both sub-tests are .57 for the HICIT sub-test and .50 for the ACE sub-test (for ages 6-9:11 only). The CA score for the HICIT ‘making inferences’ sub-test is, therefore, .07 points higher than the CA score for the inferential comprehension sub-test of the ACE. Adams et al (2001) point out that the strength of CA is related to the number of items in a sub-test. The inferential comprehension sub-test in the ACE has only nine items. The making inferences sub-test in the HICIT has 12 items, so the CA score would be expected to be slightly more reliable.

The CA scores for the idiom sub-test of the HICIT can be compared directly to the CA scores for the ACE (Adams et al, 2001) non-literal comprehension sub-test. The overall CA score for this ACE sub-section in the manual is 0.73. However, the ACE goes up to age 11:11 so the comparative CA scores for the age groups that correspond to the HICIT are given in table 4.1 below.

Table 4.1 Comparison of CA scores for the ACE non-literal comprehension sub-test and the HICIT idioms sub-test

Age group	**ACE non-literal comprehension sub-test CA**	**HICIT Idioms sub-test CA**
6:00-6:11	.63	.55
7:00-7:11	.61	.68
8:00- 8:11	.61	.73
9:00-9:11	.63	.62
CA average across the ages	**.62**	**.65**

The CA score for the HICIT idioms sub-test is, therefore, .03 points higher than the CA score for the non-literal comprehension sub-test of the ACE.

The ACE divides the 6:00-6:11 year age group into two sub-groups: 6:00-6:05 and 6:06-6:11. The younger 6 year-old group obtained a higher CA score (.7) than the older 6 year-old group (.56). The HICIT CA score for the total 6 year age group is .55, which corresponds to the ACE 6:06-6:11 age group. This suggests that the performance of 6 year-olds on non-literal comprehension tasks is inconsistent. This would fit in with Levorato and Cacciari's finding (1995, in Cain et al, 2009) that children only start to be more consistent in their understanding of idioms from the age of 7 years.

Internal consistency reliability: Rasch Analysis

Rasch analysis is a type of Item response theory (IRT). It measures the probability that an examinee answers a particular item correctly depending on the difference between the ability of the examinee and the difficulty of the item (Eckes, 2015). The analysis generated the histogram results in figures 3.5, 3.6. and 3.7.

The outliers in histogram of difficulties (figure 3.5) are from sub-section 7, situation-based emotions. The specific questions 4, 5 and 14 all test the understanding of the emotion 'happy'. Most children across all the age groups answered these questions correctly, which is why they are outliers in the data.

The questions are:

4. Lucy's dad bought her some sweets. How does she feel?
5. Alice was invited to a friend's birthday party. How does she feel?
14. Jake has won first prize in a competition. How does he feel?

The only 'happy' question not included as an outlier was number 11 (Rahal's friend has brought him a present back from his holiday. How does Rahal feel?).

The understanding of the basic emotion happy develops earliest out of all the emotions in childhood (Hadwin et al, 2015) so this is not an unexpected result. The sub-test involves the assessment of the four basic emotions: happy, sad, angry and scared. Consequently it is not appropriate to remove the three 'happy' outliers as they form part of the four elements of assessment.

Comparison of the Rasch analysis histogram of scores and the SPSS total results histogram for 6:00-9:11 year-olds (figures 3.7 and 3.8 in chapter 3) demonstrates very similar shaped histograms. Both are negatively skewed, that is there is a pile up of cases to the right and the left tail longer. The similar results from the Classical Test Theory (SPSS) analysis and the Item Response Theory (Rasch) analysis demonstrate that the item difficulty and the ability of the participant correlate well in the HICIT assessment (Yu, 2013).

4.1.2.2 Inter-rater reliability

Inter-rater reliability is the extent to which two or more assessors agree on scoring for the same population and it is a measure of consistency across testers. It is computed by two or more examiners scoring up the assessment on the same participants independently. Their test scores are then correlated (Bowling, 2009). An inter-rater reliability rating of over 75% is considered to be very good.

The inter-rater reliability in the HICIT is 98.63%. This would be considered an excellent rate of marking concordance.

Inter-rater reliability (IRR) scores given in other children's language assessments are:

- The PLAI (Blank et al, 1978): James (1986) reports that IRR fell within an acceptable range.
- The TOPS (Zachman et al, 1984): The Bernhardt et al (1990) study reports that IRR in their study was only 48.4% so it would not be appropriate to use normative data from this test.

Adams et al (2001) do not report the IRR for the ACE assessment in the manual.

Norbury and Bishop (2002) achieved 90% IRR in their inference study.

The very high IRR in the HICIT is likely to be because of the extremely detailed marking criteria provided to the markers.

4.1.2.3 Test-retest reliability

Test-retest reliability, where the test is administered twice to the same group of participants after a short space of time (eg 2-4 weeks), has not yet been carried out for the HICIT but would be a further reliability test to carry out once the final version has been produced. A positive test-retest correlation score would be over .80, as measured by Pearson's r correlation co-efficient (Adams et al, 2001).

James (1986) reported that test-retest reliability fell within the acceptable range for the PLAI (Blank et al, 1978); Carrow-Woolfolk (1999) reported scores of .92-.93 for the CASL; and Adams et al (2001) reported a score of .80 for the main test and .83 for the extended test. However, the test-retest scores for the individual sub-tests of the ACE were considerably lower. It was .47 for the inferential comprehension sub-section (the lowest rate of all the sub-tests) and .65 for the non-literal comprehension

sub-test. The low score for the inferential comprehension sub-test indicates the difficulty of reliably assessing this complex area of comprehension.
When re-testing children with the HICIT one might expect to get the same answers each time in the sub-sections that have only single acceptable responses (sub-tests 7, 8 and 9). The remaining sub-tests have multiple acceptable responses (either single word: sub-sections 1 and 11 or phrase or sentence level: sub-sections 2, 3, 4, 5, 6, 10 and 12). In these sub-sections it would be expected that many of the children's responses could be different each time of testing. However, the very detailed marking criteria for acceptable answers for the HICIT should help to keep the test-retest reliability score relatively high (ideally higher than the equivalent sub-test in the ACE of .47).

4.2 Primary study aim 2 and related hypothesis

To provide normative and statistically significant data for inferential and idiom comprehension in typically developing children aged 5:00 to 9:11.

Related Hypothesis: Alternate Hypothesis (H1): The older age groups will have higher inferential and idiom comprehension scores.

4.2.1 Descriptive statistics. Total test analysis

Descriptive statistics were completed for the 6:00-9:11 cohort (200 children) as they were assessed on all 12 sub-sections of the final version of the HICIT. The results demonstrated that the mean score, out of a possible maximum score of 140 increased with age: 88 in the 6:00-6:11 age group, 103 in the 7:00 to 7:11 age group, 107 in the 8:00 to 8:11 age group, and 120 in the 9:00 to 9:11 age group This indicates that 6:00 to 9:11 year-old children's inferential and idiom comprehension improves with age, fulfilling the first aim of providing normative data for these areas of comprehension.

The range was biggest in the 6:00 to 6:11 age group (51-123, a range of 72). The other age groups had ranges of 40, 41 and 34 respectively. The explanation for such a wide range in the youngest age group is that this group included the two extreme outliers, one with a much higher than average score for that year-group (123/140) and one with a much lower than average score (51/140). If these two extreme outliers are removed from the calculation the range becomes 62-108 (a range of 46 which is similar to the range in the other groups).

The Kolmogorov-Smirnov Significance value was .2 for each age group, demonstrating that there is a normal distribution within each age band.

<u>Sub-sections analysis</u>

Normative data were also produced for each sub-section of the HICIT. This included the age group 5:00-5:11 for the sub-tests 1 to 8 so the age range was 5:00 to 9:11 and the sample size was 250 for these.

All of the mean sub-test scores increased with age apart from the Strange Stories sub-section, where the mean score for 7:00-7:11 year-olds (9.86) was .02 greater than the mean score for 8:00-8:11 year-olds (9.84).

The majority of the distributions in each age groups in each sub-section were negatively skewed (ie overall more children answer the questions correctly than incorrectly). The only significantly positively skewed sub-section age groups (indicating that overall more children answer the questions incorrectly than correctly) were:

- Situation-based emotions, age group 6:00-6:11
- Faux pas, age groups 6:00-6:11 and 8:00-8:11
- Idioms, all age groups except 9:00-9:11

The only sub-section age bands with normal or near normal distributions were:

- Deductions, age group 6:00-6:11 (slightly positively skewed)
- Why questions, age group 5:00-5:11 (negatively skewed)
- Making inferences, age group 5:00-5:11 (positively skewed)
- Short passage inferences, age group 5:00-5:11
- Idioms, age group 7:00-7:11 (positively skewed) and age group 9:00-9:11 (negatively skewed)

<u>Strange Stories section</u>

It became apparent during the testing process that question 4a in the Contrary Emotions story can be correctly answered with both a Yes or No question (‘Yes’ as James doesn’t want to go swimming because bullies are at the pool, ‘No’ as he would really like to go swimming but he won’t because bullies are at the pool). Consequently both answers were accepted thus effectively removing this question from the results. However 4b (the justification for their answer in 4a) is a useful question to ask so both questions will be maintained in the final test.

Faux Pas sub-section

The scores in this subsection out of a possible maximum of 12 are:

Table 4.2 Faux pas mean scores in the HICIT

Age	Mean
6-6:11	7.2
7-7:11	8.54
8-8:11	8.84
9-9:11	10.08

Baron-Cohen et al's (1999) Faux Pas study results (out of a possible maximum of 10) were:

Table 4.3 Faux pas mean scores in Baron-Cohen et al's (1999) study

Age	Sex	Mean
7	Boys	2.9
7	Girls	3.8
7	Combined	3.4
9	Boys	4.6
9	Girls	7.2
9	Combined	5.9

It is difficult to make a direct comparison between the two sets of results as, in the Baron-Cohen (1999) study, the children had to answer four questions correctly (faux pas detection, identification, comprehension, and false belief) to get a score of 1. In contrast, in the HICIT Faux pas sub-test the children were credited with 1 point for every question (out of 3) they answered for each faux pas scenario (there were 4 of these, so the maximum possible score was 12).

All of the above sections reached almost ceiling scores by 9:11. This indicates that inferential comprehension, as assessed by the HICIT, is nearly completely developed by 10 years of age.

Idiom sub-section

This is the only sub-section of the HICIT where no child received a ceiling score of 20 by the age of 9:11 and where the mean score of the oldest children was only just over half of the maximum possible score. This was to be expected, as the literature on idiom comprehension in typically developing children reports that it continues into adolescence and even adulthood (Nippold and Duthie, 2003, Cain et al, 2009, Lundblom and Woods, 2012).

A student from Manchester Metropolitan University (MMU) subsequently carried out an extension study of the HICIT idiom comprehension sub-test for her final year research project (Maddocks, 2015). She assessed 25 girls and 25 boys in two further age groups (10:00-10:11 and 11:00-11:11) with the sub-test. Fifty children per year-group is the same as the number for the initial HICIT study so allowed extension figures for the idiom test to be generated.

The combined results from the original HICIT and the extension idiom study are given in table 4.4 below.

Table 4.4 Results for the HICIT Idioms sub-section by age and gender

Age Band	Sex	Mean (/20)	Range
6:00-6:11	Male	3.64	0-10
6:00-6:11	Female	2.68	0-8
7:00-7:11	Male	5.04	0-11
7:00-7:11	Female	6.44	0-10
8:00-8:11	Male	6.48	0-12
8:00-8:11	Female	7.64	3-13
9:00-9:11	Male	11.28	5-18
9:00-9:11	Female	11.52	5-17
10:00-10:11	Male	11.08	4-16
10:00-10:11	Female	13.68	9-19
11:00-11:11	Male	13.56	4-17
11:00-11:11	Female	14.96	9-20

As can be seen, idiom comprehension continues to increase with age and is still not at ceiling level on this sub-test by nearly 12 years of age.

Maddocks (2015) found a significant difference between the age groups in this sub-test.

These results demonstrate a clear developmental progression in idiom comprehension from 6:00 up to 11:11 which corresponds well to the research findings in this area (Nippold and Taylor, 2002, Caillies and Le Sourn-Bissaoui, 2008, Whyte et al, 2014). The sub-test had still not reached ceiling level by 11:11 which also supports previous research that idiom comprehension continues to develop throughout adolescence and into adulthood (Cain et al, 2009, Conner et al, 2011).

As the HICIT uses the most familiar idioms and does not sub-categorise them into opaque or transparent, findings from the literature in this area could not be applied to the results. The preferred method in the research for assessing idiom comprehension is a multiple choice test of idioms in context (Cain et al, 2009). The HICIT idiom comprehension assessment is a definition task, and it assesses idioms out of context.

The current researcher followed Norbury's (2004) lead on this format, as Norbury found that multiple choice tests with foils made the task too easy and a definition task gave a good insight into the child's thought processes when trying to define the idioms. In addition, the Test of Language Competence (Wiig and Secord, 1989) uses a definition task for its level 1 (5:00-9:11) idiom sub-test. In the methodology chapter (2), the current researcher criticised the Test for Language Competence (Wiig and Secord, 1989) for having idioms as the acceptable answers for their idiom definitions. The HICIT marking criteria also include some idiomatic definitions of the idioms but the difference is that it also includes non-idiomatic definitions. The HICIT does not include any trial items for idioms but it does have a lead in example and explanation of the idiom 'give me a hand' before the definition task commences.

4.2.2 Inferential statistics

Aim 2. Related Hypothesis - Alternate Hypothesis (H1): The older age groups will have higher inferential and idiom comprehension scores.

The results chapter (3) reported on the two-way ANOVA looking at the effects of age group and gender on the total test score. The results demonstrated a significant effect for age group (.00: F=71.52)

This supports the alternate hypothesis: that the older age groups have higher inferential and idiom comprehension scores.

The developmental progression of idioms has already been discussed in relation to the relevant literature in the above section. The finding that inferential comprehension develops with age is supported by the previous studies carried out by Letts and Leinonen (2001), Botting and Adams (2005), Bernicot et al (2007), Silva and Cain (2015) and Filiatrault-Veilleux et al (2015).

4.3 Secondary study hypotheses

3.5.1 Secondary hypothesis 1: Gender effects

Gender effects from the HICIT total test scores

The results chapter (3) reported on the two-way ANOVA looking at the effects of age group and gender on the total test score. No significant effect for gender was found (.21: F=1.60)

This supports the null hypothesis that the gender of the participant has no effect on inferential and idiom comprehension scores.

Curtis and Foord (1996) found no significant effect for gender when they administered the PLAI (Blank et al, 1978) and the Peabody Picture Vocabulary Test-Revised (Dunn and Dunn, 1981) to 43 typically developing Australian children aged 4:06-4:11.

The current study's findings of no gender effect are contrary to Bornstein's (2004) study which found that girls perform better than boys in language in the early years. However, Bornstein's children were aged between 1 and 6:10 and the norms for the main part of the HICIT are from 6:00 onwards, so the same findings are not likely to apply. The findings also do not support the opposite view, expressed by Lovas (2011), that boys have better verbal comprehension and verbal reasoning skills than girls until adolescence.

Gender effects in sub-tests of the HICIT

Idiom comprehension

Maddocks (2015) found a significant effect for gender in her idiom extension study (boys and girls aged 10:00 to 11:11 years), with girls performing better than boys. The current researcher found the opposite in the HICIT study with the boys aged 6:00- 9:11 performing slightly better in idiom comprehension in comparison to the girls. However, there was no statistically significant difference in total test scores between boys and girls in the HICIT. There is little previous research into gender differences in idiom comprehension and other idiom studies (Nippold and Duthie, 2003, Kana et al, 2012) have not found a significant gender difference.

Faux pas comprehension

Baron-Cohen et al's (1999) Faux Pas study results (out of a possible maximum of 10) are given below:

Table 4.5 Baron-Cohen et al's (1999) Faux Pas study results

Age	**Sex**	**Mean**
7	Boys	2.9
7	Girls	3.8
7	Combined	3.4
9	Boys	4.6
9	Girls	7.2
9	Combined	5.9

The girls performed better than the boys in this task which is what the researchers had predicted since there is considerable evidence that females are more socially developed than their male peers. As there was no significant difference for gender in the total HICIT scores, the boys and girls scores in the faux pas sub-test were not analysed separately, so a comparison with Baron-Cohen's (1999) faux pas gender results is not possible.

4.3.2. Secondary hypothesis 2 : Socio-economic effects

The one-way ANOVA results, reported in the results chapter (3), demonstrated that there was a significant difference in the total test scores across socio-economic groups, with the higher socio-economic groups performing better than the lower socio-economic groups.

This supports the alternate hypothesis that the socio-economic status of the participant will have an effect on inferential and idiom comprehension scores.

This fits in with Curtis and Foord's (1996) finding that 4:06 to 4:11 year-old Australian children from the centres in the lower socioeconomic areas obtained lower mean scores on two language assessments than the children from the centres in higher socioeconomic areas. It also corresponds to Van Kleeck's (2008) reports that children from lower socioeconomic backgrounds experience more difficulty with inferential comprehension both verbally and in reading. Many other studies have linked lower socio-economic status to lower performance on standardised language tests (Spencer et al, 2016).

However, the initial conclusion drawn from the one-way ANOVA is not a robust one, as the age bands assessed in the different socio-economic areas in the HICIT study were not the same as each other. Post-hoc tests were carried out on the data as there was a significant main effect for socio-economic status.

The cohorts that could be compared are summarised in the results chapter, section 3.3.3.1. Cohort 1 consisted of children aged 6:00 to 9:11. 60 children attended schools that had a FSM % of 9.6 and 36 attended schools that had a FSM % of 53.6. An independent samples T-test was carried out on these data and demonstrated a significant difference (.00) between the two groups, with the higher socio-economic status group performing better than the lower socio-economic status group This result initially appears to support the main study's Hypothesis 3. However, there is a confounding variable in these results. Namely that the mean age of the children in the lower socio-economic status groups is 6 months less than the mean age of the children in the higher socio-economic group. Since like groups have not been compared, the T-test result is inconclusive.

Cohort 2 consisted of children aged 7:00-7:11. Fourteen children had a FSM % of 9.8% and ten had a FSM % of 50.3%. An independent T-test result carried out on this data demonstrated no statistically significant difference (.06) between the two groups. This contradicts the findings from the main study and from group 1. However, the confounding variable in this case is that the sample sizes in both groups is extremely small, making the power of the T-test very low and the result not very reliable. For more conclusive results, more testing of the HICIT with matched age groups in schools with different socio-economic status is needed to see if this is a variable that needs to be taken into account when calculating the test normative data.

4.4 Qualitative analysis of the data

The main results of this study have been quantitative. However, qualitative results are also important, particularly when analysing the responses of children with communication impairments.

In the Test of Language Competence (Wiig and Secord, 1989) the possible choices in the idiom sub-test are qualitatively sub-categorised into: nonrelated (N), literal (L), opposite (O) or matching (M) response. In her idiom extension study, which included the idiom data from the HICIT study, Maddocks (2015) used the categories: 'Idiomatic', 'Literal', 'Don't know' and 'Other' in her qualitative analysis of the typically developing children's responses. She found that the children's literal answers declined as they got older. Six year-old children gave a literal definition of an idiom 27.5% of the time whereas eleven year-old children only gave literal definitions 1.5% of the time. This fits in with the developmental data on typical development of idiom comprehension and Huber-Okrainec et al's (2005) findings that 6 year-olds make more literal interpretations of idioms than older children.

In their study the inferential comprehension of language-impaired and typically developing children Letts and Leinonen (2001) saw a developmental progression in the typically developing children. The questions involving feelings were answered by the younger children using the generic terms (sad, happy etc). The older children used a more diverse emotional vocabulary. The results from the HICIT development study support this in terms of both comprehension and expression of emotions. In section 7, situation-based emotions, the younger children selected ‘sad’ as the answer for many of the ‘angry’ and ‘scared’ questions whereas the older children had no difficulty identifying all four emotions. In section 11, faux pas, the younger children used emotional vocabulary such as sad, upset, whereas the older children used terms such as annoyed, disappointed, furious, shocked, offended, grumpy, depressed, guilty, embarrassed, regretful, terrible, full of shame. The younger children in the HICIT study sometimes gave impractical solutions to some of the problems. For example suggesting the use of a ladder or helicopter to get up to the top floor of a block of flats; suggesting going by bike, car or plane to talk to someone who lives 100s of miles away, ‘the chicken ran away’ as the reason for the family not having chicken for dinner. The younger children also made more errors with the questions requiring well developed theory of mind skills, in particular the ‘How do you know that X?’ questions. Some of them answered these as if they were ‘Why X’ or ‘What could you do if X?’ questions. This is an important finding, as Child C above made many errors in this section so his difficulties here could be explained by his generally delayed language comprehension whereas his other errors would be described as ‘unusual’ and a possible feature of his pragmatic difficulties.

In their inference study with children with communication impairments and typically developing children, Norbury and Bishop (2002) reported that ‘wrong inference’ was the most common error type. ‘Odd responses’ were very rare and no typically developing children gave these. Letts and Leinonen (2001) described any kind of incorrect responses as ‘problematical responses’. They reported that these diminished with age. The 8 year-olds in their study gave a higher number of ‘don’t know’ responses. They interpreted this as evidence that the children became more aware of the difficulties associated with giving a logical justification for answers at around this age. In the HICIT assessment many younger children used ‘don’t know’ responses. Contrary to Norbury and Bishop’s (2002) findings a small number of typically developing children gave ‘odd’ responses to questions. These are labelled as ‘unusual’ responses by the current researcher. See the end of the results chapter (chapter 3) and appendix xvii for multiple examples of these. It can be seen from these examples that it is predominantly the younger children (up to 6:11) who give these types of responses. It is important to be aware of this fact when qualitatively analysing the errors of children with communication impairments as some of the errors that could be attributed to a specific client group also occur in typically developing children. For example some of the unusual answers given by the typically developing children at the end of chapter 3 are very similar to some of the ‘unusual’ responses from the children with HFA and diGeorge Syndrome described above. Sensitivity and specificity (measures of classification accuracy) need to be taken into account when scoring the HICIT assessment to prevent over- or under-diagnosis of inferential and idiom comprehension difficulties. Sensitivity is a measure of the correct classification of true positives and specificity is a measure of the correct classification of true negatives. High values in both measures are indicators of very accurate classification of the test questions (Hatcher, 2013). The fact that typical children make some unusual errors in their test responses should not affect the specificity and sensitivity of the HICIT as such errors occur very infrequently with this population. In addition the answers would be marked as incorrect and would therefore be accounted for in the descriptive statistics.

Adams et al (2009) recommend doing an analysis of the error types in inferential comprehension for different sub-groups of children with communication impairments. The HICIT lends itself extremely well to this sort of analysis and more detailed qualitative analysis of the results from the normative sample would help clinicians when assessing children with comprehension difficulties to make judgements about what features are delayed typical development and what features indicate specific communication impairments.

4.5 Discussion of the exploratory case studies findings

These results of the four exploratory case studies reported in the Results chapter, section 3.5, lend further support to the findings from the literature that children with ASD and pragmatic impairments experience considerable difficulty with inferential comprehension of verbal language (Leinonen and Letts, 1997, Botting and Adams, 2005, Adams et al, 2009, Lucas and Norbury, 2015). They also lend support to previous research findings that children with ASD and pragmatic impairments have particular difficulty understanding non-literal language such as idioms (Rinaldi, 2000, Dennis et al, 2001, Whyte et al, 2014). The results also support the findings from previous studies that children with ASD have particular difficulty understanding social inferences that depend on a good Theory of Mind (TOM) ability (Happe, 1994, Jolliffe and Baron-Cohen, 1999, Kaland et al, 2005, O'Hare et al, 2009, Baron-Cohen et al 1999, Pearson and Pillow, 2016).

4.6 Overall evaluation of the HICIT

The limitations of the HICIT development process and final assessment are given below:

Limitations

- The small number of children in the pilot study (10 children per year group) meant that decisions about which items to retain and remove from the original HICIT were based on less reliable statistical analyses.
- The standardisation sample size of the final HICIT is not as big as some of the existing British standardised language assessments (eg the ACE, Adams et al ,2001).
- The internal consistency reliability scores (Cronbach's Alpha) are not as good as some of the American language assessments.
- The sample did not have equal numbers of children from different socio-economic areas, making the examination of the effect of socio-economic status on inferential and idiom comprehension inconclusive.
- Data was not collected on the ethnicity of the children taking part in the study so this aspect could not be examined. However, in the ACE (Adams et al, 2001) standardisation process 93.1% of the standardisation sample was white (the percentage in England and Wales when the ACE was published was 94.1%). The population tested by the current researcher would have a similar percentage, but she does not have the ethnicity information for the assessments carried out by the students in other parts of the country.

Advantages of the final HICIT:

- It is a standardised test.
- The standardisation sample is British and British English is used for the test questions.
- The sample size (50 per year group, 250 in total) is reasonable for a published test and excellent for a part-time PhD study.
- The Cronbach's Alpha (CA) reliability scores are slightly better than those for the relevant sub-tests of a British language assessment (the ACE, Adams et al, 2001). The numbers of questions in each sub-section are relatively small (6 to 12, with one section of 20) which lowers the CA figure considerably.
- The inter-rater reliability score (98.63%) is excellent.
- The very clear, detailed marking criteria improve the internal reliability of the test.
- Unlimited repetitions of the test questions are allowed to reduce any demands on verbal memory.
- No other standardised test of inferential comprehension and idioms has as many questions in total as the HICIT. The 5:00-5:11 year-old children are asked 84 inferential comprehension questions. The inferential comprehension part of the HICIT 6:00 to 6:11 has 120 questions and the idiom section has 20 questions. The comparable sub-tests in the ACE have 9 and 15 questions per section respectively.
- Normed scores will be available for individual sub-tests as well as the overall tests so selected sub-tests can be administered, so cutting down test administration time.
- The assessment contains questions which assess social cognition, in particular theory of mind, which is a recommendation from previous research in this area (Adams et al, 2009).

4.7 Further amendments to, analysis of and tests to carry out with the HICIT.

Suggested amendments to the HICIT

The following amendments are justified in the above discussion of the results:

- Sub-test 8 (Belief-based emotions) should be removed from the assessment as the mean only increased very slightly over the whole population (from 5.18 at 5:00-5:11 years to 6.94 at 6:00-6:11, out of a possible maximum score of 8). There are only two possible answers to these questions (happy or sad), so allowing a 50% chance of answering correctly. This is the most likely explanation for the similarity of scores across the whole age range.
- Sub-tests 3 (Predictions) and 4 (Formulating Solutions) reached near to their mean ceiling score by 7:11. These sub-tests could be removed for the older children and just be administered to the 5:00- 7:11 age group
- Sub-tests 6 (Making inferences from short passages), 9 (Mental state verbs), 10 (Strange Stories) and 11 faux pas all have very similar mean scores for the 7:00-7:11 and the 8:00-8:11 age bands. These sub-test means could therefore be combined into a joint age group of 7:00-8:11.
- Sub-test 12 (Idioms). The extended norms up to 11;11 from the Maddocks (2015) study can be added in to the results.

Acting upon the fact that only sub-tests 1-8 were administered to the 5 year-olds and implementing the above changes would create four separate total tests:

1. **5:00-5:11**
2. **6:00-7:11**
3. **8:00-9:11**
4. **Idioms Test 6:00-11:11**

The scoring will need to be amended to reflect these changes. An advantage of these amendments is that the shortened tests will reduce the time taken to administer the complete test (currently 30 minutes if the responses are audio-recorded and transcribed later and 40-45 minutes if the responses are transcribed in situ).

As well as the age equivalents and range of scores, standard scores, percentile ranks and confidence intervals will need to be generated for the test.

<u>Further analysis of the standardisation sample HICIT results</u>

A more detailed qualitative analysis of the HICIT responses of the 250 children in the study could be done to extract more information on typical developmental trends in inferential and idiom comprehension.

A gender analysis of specific sub-tests could be carried out, particularly those involving social awareness and empathy as researchers have found that girls perform better than boys in these.

4.8 Future tests and research studies needed

Validity and Reliability tests to implement

Tests of concurrent validity (comparing the results of the HICIT to the results of other similar language assessments with the same children) and of test-retest reliability (re-administering the HICIT to the same children after a short time gap and comparing the results).

Socio-economic status

As the results on whether or not socio-economic status affected the HICIT scores were inconclusive, a further large scale study (possibly as part of the concurrent validity study) with equal groups from different socio-economic backgrounds could be conducted. Then any statistical analyses would be more robust and conclusive. Ethnicity would also be recorded in this study to see if this significantly affects assessment results.

Large scale studies using the HICIT with children with communication impairments

Large scale studies using the HICIT with children with receptive language disorders, pragmatic language impairment, high functioning autism, other developmental disabilities which include comprehension difficulties. Quantitative and qualitative analysis of the results would demonstrate if the HICIT will be a useful tool to aid differential diagnosis. In view of the research findings into the correlations between different areas of development (Oakhill and Cain, 2012, Silva and Cain, 2015, Lucas and Norbury, 2015) simultaneous assessment of the children's cognitive skills, theory of mind, verbal memory, comprehension of grammatical structures, expressive syntax and receptive and expressive vocabulary would help to determine what the primary cause of inferential comprehension difficulties is.
Longitudinal studies of children receiving intervention for inferential comprehension difficulties who are assessed at yearly intervals with the HICIT could be useful to

help document the developmental trajectory of inferential and idiom comprehension and demonstrate if intervention is being effective.

Acquired brain injury studies

The HICIT could be used as a screen to assess inferential and idiom comprehension in children and adults with acquired brain injuries.

4.9 Feedback obtained on the potential clinical usefulness of the HICIT

The current researcher presented her PhD findings so far at the North West Network Conference, attended by Speech and Language Therapists (SLTs) and teachers specialising in working with communication impaired children, in June 2015. See appendix ixx for a copy of the power point presentation.

She also presented her findings at the Liverpool speech and language therapy team day in December 2015.

The feedback from both of these presentations was extremely positive, with SLTs expressing their views that the HICIT would be a very useful clinical tool. Many participants asked for the test to be available to them as soon as possible.

4.10 Getting the HICIT published

The ultimate aim of this study was to create a clinically useful, standardised, British test of inferential and idiom comprehension for children aged 5:00 to 9:11. The ideal scenario would be to get the test published. Four exploratory studies carried out with children with communication impairments and feedback from practising clinicians give early indications that the test is potentially very clinically useful to help with the differential diagnosis of different types of comprehension disorders.

To get the test published (once the suggested amendments have been made) the following procedures would need to be followed:

- Obtain permission from any researchers whose ideas have been incorporated into the HICIT to use their intellectual property.
- Approach the test company Pearson's to see if they would be interested in publishing the test. They formerly published Wendy Rinaldi's (1996) Understanding Ambiguity assessment which was standardised on only 15 typically developing children per year group, so the moderate standardisation

sample should not be an issue. However, this was 20 years ago so their standardisation criteria are likely to be more rigorous now.

- Approach the Black Sheep Press company.
- Publish it in house in Manchester Metropolitan University (MMU).
- Discussion with MMU will be required as to how the intellectual property rights of the HICIT will be shared between the author and the University.

4.11 Concluding statements

Bernhardt (1990) states that speech and language therapists have been searching for years for a tool to assess children's verbal reasoning skills. Norbury and Bishop (2002) emphasize the importance of inferential comprehension for successful functional communication and academic success. Dockrell and Messer (1999) point out that there are very few tools to assess inferencing skills to help work out the implications for intervention. Ryder et al (2008) highlight the importance of being able to assess inferential comprehension in depth, to be able to provide the most appropriate intervention and educational support. Filiatrault-Veilleux (2015) and Bishop et al (2009) describe how inferential comprehension is being increasingly targeted in speech and language therapy. They add that until we have more information on typical development in this area and an effective and reliable assessment we will not be able to target intervention or measure its outcome appropriately.

The HICIT fulfils most of these identified needs from the literature. The main aim of the study, to create a robust, clinically useful, standardised British test of inferential and idiom comprehension for children aged 5:00 to 9:11 has been achieved.

References

Adams, C., Cooke, R., Crutchley, A., Hesketh, A. and Reeves, D. 2001. ACE – Assessment of Comprehension and Expression 6-11. Windsor: NFER-Nelson.

Adams, C. (2002) 'Practitioner review: The assessment of language pragmatics.', *Journal of Child Psychology and Psychiatry and Allied Disciplines.,* 43(8), pp. 973-987.

Adams, C., Clarke, E. and Haynes, R. (2009) 'Inference and sentence comprehension in children with specific or pragmatic language impairments.', *International Journal of Language and Communication Disorders.,* 44(3), pp. 301-318.

Andersson, L. (2005) 'Determining the adequacy of tests of children's language.' *Communication Disorders Quarterly,* 26(4) pp. 207.

APA (1994). Diagnostic and statistical manual of mental disorders. 4th Edition. Arlington, VA.: American Psychiatric Publishing.

APA (2013) *DSM5* www.dsm5.org (Accessed: 20/02 2016).

Armitage, R. (1994) *The Lighthouse Keeeper's Lunch.* London.: Scholastic.

Baron-Cohen. S., Leslie, A.M. and Frith, U. (1985) 'Does the autistic child have a 'theory of mind?' *Cognition,* 21 (1) pp. 37-46.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R. and Plaisted, K. (1999) 'A new test of social sensitivity: Detection of faux pas in normal children and children with Asperger's Syndrome.', *Journal of Autism and Developmental Disorders,* 29, pp. 407-418.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V. and Lawson, J. (2001) 'Are intuitive physics and intuitive psychology independent? A test with children with Asperger syndrome.', *Journal of Developmental and Learning Disorders.,* 5, pp. 47-78.

Baron-Cohen, S., Tager-Flusberg, H. and Lombardo, M., V. (Eds). (2013) *Understanding other minds: Perspectives from developmental cognitive neuroscience.* 3rd Edition. Oxford: Oxford University Press.

Barnes, M. A. and Dennis, M. (1998) 'Discourse after early-onset hydrocephalus: Core deficits in children of average intelligence.', *Brain and language.,* 61(3), pp. 309-334.

Barnes, M. A. and Dennis, M. (2001) 'Knowledge-based inferencing after childhood head injury.', *Brain and language.,* 76(3), pp. 253-265.

Barrett, M., Huisingh, R., Zachman, L., Blagden, C. and Orman, J. 1992. The Listening Test: Linguisystems Inc.

Basit, T.N, Hughes, A., Iqbal, Z and Cooper, J. (2015) 'The influence of socio-economic status and ethnicity on speech and language development', *International Journa lof Early Years Education,* 23 (1), pp.115-133

Bernhardt, B. (1990) 'A Test of Problem Solving (TOPS).', *Language, Speech and Hearing Services in Schools.,* 21, pp. 98-101.

Bernicot, J., Laval, V. and Chaminaud, S. (2007) 'Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics?' *Journal of Pragmatics*, 39(12) pp. 2115-2132.

Bishop, D. V. M. 1989. Test for Reception of Grammar (second edition). Manchester UK: Age and Cognitive Performance Research Centre. University of Manchester.

Bishop, D. V. M. and Adams, C. (1992) 'Comprehension problems in children with specific language impairment: literal and inferential meaning.', *Journal of speech and hearing research.,* 35, pp. 119-129.

Bishop, D., V. M. (1997) *Uncommon understanding. Development and disorders of language comprehension in children.* East Sussex: Psychology Press Ltd.

Bishop, D. 2003. Test for Reception of Grammar (TROG-2). Oxford, UK.: Pearson Assessment.

Bishop, D. 2003. The Children's Communication Checklist - 2nd Edition: The Psychological Corporation.

Bishop, D. V. M. 2004. Expression, Reception and Recall of Narrative Instrument (ERRNI). London: Psychological Corporation.

Bishop, D.V.M., Snowling, M.J., Thompson, P.A., Greenhalgh, T and CATALISE Consortium (2016). CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development. Phase 2: Terminology. PeerJ.doi:10.7287/peerj.preprints.2484v1

Black Sheep Press resources: www.blacksheeppress.co.uk

Blank, M., Rose, S. and Berlin, L. (1978a) *The Language of Learning in the Preschool Years.* New York: Grune & Stratton.

Blank, M., Rose, S. and Berlin, L. (1978b) *The Language of Learning in Practice.* New York: Grune & Stratton.

Blank, M., Rose, S. A. and Berlin, L. J. 1978. Preschool Language Assessment Instrument (PLAI). 1st. ed.: ProED

Blank, M., Rose, S. A. and Berlin, L. J. 2003. Preschool Language Assessment Instrument (PLAI). 2nd. ed.: Linguisystems Inc.

Botting, N. and Adams, C. (2005) 'Semantic and inferencing abilities in children with communication disorders.', *International Journal of Language and Communication Disorders.,* 40(1), pp. 49-66.

Bornstein, M. H., Chun-Shin, H. and Haynes, O. M. (2004) 'Specific and general language performance across early childhood: stability and gender considerations.', *First Language,* 24(3), pp. 267-304.

Boucher, J. (1998) 'SPD as a distinct diagnostic entity: logical considerations and directions for future research.' *International Journal of Language and Communication Disorders,* 33 (1) pp.71-81

Bowers, L., Huisingh, R., Barrett, M., Orman, J. and LoGiudice, C. 1994. Test of Problem Solving - Elementary, Revised. East Moline, IL.: LinguiSystems.

Bowers, L., Huisingh, R. and LoGiud ice, C. 2007. Test of Problem Solving 2 - Adolescent (TOPS-2 Adolescent). Linguisystems.

Bowling, A. (2009) *Research methods in health. Investigating health and health services.* 3rd. ed., New York.: Open University Press.

BPS (2014) *Code of Human Research Ethics [Online] [Accessed on 20th February 2016].*http://www.bps.org.uk/sites/default/files/documents/code_of_human_research_ethics.pdf

Bracken, B.A 2002. The Bracken School Readiness Assessment. London: Pearson Education inc.

Brandel, J., Gillam, R., B., Hoffman, L., M. and Loeb, D. F. (2011) 'Concurrent and construct validity of oral language measures with school-age children with specific language impairment.', *Journal of Speech, Language and Hearing Research,* 54(6), pp. 1597.

Brugha T, Cooper SA, McManus S, Purdon S, Smith J, Scott FJ, Spiers N, Tyrer F (2012) *Estimating the Prevalence of Autism Spectrum Conditions in Adults: Extending the 2007 Adult Psychiatric Morbidity Survey.* The Health and Social Care Information Centre, NHS

Bryan, K. L. 1995. The Right Hemisphere Language Battery (2nd Edition). London.: Whurr Publishers Ltd.

Burdon. P., Dipper, L. and Cocks, N. (2016) Exploration of older and younger British adults' performance on The Awareness of Social Inference Test

(TASIT) *International Journal of Language and Communication,* 51(5), pp.589-593

Caillies, S. and Le Sourn-Bissaoui, S. (2008) 'Children's understanding of idioms and theory of mind development.', *Developmental Science,* 11(5), pp. 703-711.

Cain, K. and Oakhill, J. V. (1999) 'Inference making ability and its relation to comprehension failure in young children.', *Reading and writing.,* 11(5-6), pp. 489-503.

Cain, K., Oakhill, J. V., Barnes, M. A. and Bryant, P. E. (2001) 'Comprehension skill, inference-making ability and their relation to knowledge.', *Memory & Cognition.,* 29(6), pp. 850-859.

Cain, K., Towse, A. S. and Knight, R. S. (2009) 'The development of idiom comprehension: An investigation of semantic and contextual processing skills.', *Journal of Experimental Child Psychology,* 102, pp. 280-298.

Carrow-Woolfolk, E. 1999. Comprehensive Assessment of Spoken Language (CASL). Pro-Ed.

Channon, S., Pellijeff, A. and Rule, A. (2005) 'Social cognition after head injury: Sarcasm and theory of mind.', *Brain and Language.,* 93, pp. 123-134.

Christie, J., Clark, W. and Mortensen, L. 1986. Mount Wilga High Level Language Test. Hornsby, Australia: Speech Pathology Department, Mount Wilga Rehabilitation Centre.

Conner, P., Hyun, J., Wells, B., O'Connor., Anema, I., Goral, M., Monereau-Merry, M.-M., Rubino, D., Kuckuk, R. and Obler, L., K. (2011) 'Age-related differences in idiom production in adulthood.', *Clinical linguistics and phonetics.,* 25(10), pp. 899-912.

Crais, E. R. and Chapman, R. S. (1987) 'Story recall and inferencing skills in language/learning disabled and nondisabled children.', *Journal of speech and hearing disorders.,* 52, pp. 50-55.

Crespo, N., Manghi, D., Garcia, G. and Caceres, P. (2007) 'Attention deficit and understanding of non-literal meaning: The interpretation of indirect speech acts and idioms.', *Revista de Neurologia.,* 44(2), pp. 75-80.

Cresswell, J., W. (2014) *Research design. Qualitative, Quantitative and Mixed Methods Approaches.* 4th international student Edition. London: Sage.

Crotty, M. (1998) *The foundations of social research. Meaning and perspective in the research process.* London: Sage.

Crystal, D. (2008) *A dictionary of linguistcs and phonetics.* 6th Edition. Oxford.: Blackwell.

Culatta, P., Page, J. L. and Ellis, J. (1983) 'Communicative performance screening school.', *Journal of Speech and Hearing Services in Schools.,* 14, pp. 6-74.

Curtis, J. and Foord, W. (1996) 'Performance of Australian 4.06 to 4.11 Year Olds on the Preschool Language Assessment Instrument.', *Australian Communication Quarterly.,* Autumn.

Dennis, M. and Barnes, M. A. (1993) 'Oral discourse after early-onset hydrocephalus - linguistic ambiguity, figurative lnguage, speech acts, and script-based inferences.', *Journal of Pediatric Psychology.,* 18(5), pp. 639-652.

Dennis, M. and Barnes, M. A. (2001) 'Comparison of literal, inferential and intentional text comprehension in children with mild or severe closed head injury.', *Journal of head trauma rehabilitation.,* 16(5), pp. 456-468.

Dennis, M., Lazenby, A., L. and Lockyer, L. (2001) 'Inferential language in high-functioning children with autism.', *Journal of Autism and Developmental Disorders.,* 31(1), pp. 47-54.

DfE (2012) *Statistical First Release. Schools, pupils and their characteristics.*

Dockrell, J. and Messer, D. (1999) *Children's language and communication difficulties. Understanding, identification and intervention.* London: Continuum.

Dodwell, K. and Bavin, E. L. (2008) 'Children with specific language impairment: an investigation of their narrative and memory.', *International Journal of Language & Communication Disorders.,* 43(2), pp. 201-218.

Dollaghan, C.A (2007): *The handbook for evidence-based practice in communication disorders.* Baltimore, MD, Paul H. Brookes.

Dunn, L. M. and Dunn, L. M. 1981. Peabody Picture Vocabulary Test - Revised. Circle Pines, MN: American Guidance Service.

Dunn, L. M., Whetton, C. and Burley, J. 1998. British Picture Vocabulary Scale II. Windsor.: NFER-Nelson.

Eckes, T. (2015) *Introduction to many-facet Rasch measurement. Language testing and Evaluation.* 2nd Edition. Oxford: Peter Lang GmbH InternationalerVerlag der Wissenschaften.

Edwards, S., Letts, C. and Sinka, I. 2011. The New Reynell Developmental Language Scales. 4th Edition. (NRDLS). London: GL Assessment.

Eisele, J. A., Lust, B. and Aram, D. M. (1998) 'Presupposition and implication of truth: Linguistic deficits following early brain lesions.', *Brain and Languge.,* 61(3), pp. 376-394.

Elks, E. and McLachlan, H. 2004. The Test of Abstract Language Comprehension. Cornwall.: Elklan Publications.

Elks, E. and McLachlan, H. 2010. The Test of Abstract Language 2. Cornwall: Elklan Publications.

Elks, E. and McLachlan, H. (2005) *Language Builders.* Cornwall.: Elklan.

Feast, L. and Melles, G. (2010) *Epistemological positions in design research. A brief review of the literature.* University of New South Wales, Sydney, Australia.:

Ferstl, E. C., Guthke, T. and von Cramon, D. Y. (2002) 'Text comprehension after brain injury: Left prefrontal lesions affect inference processes.', *Neuropsycholgy.,* 16(3), pp. 292-308.

Ferstl, E. C., Walther, K., Guthke, T. and von Cramon, D. Y. (2005) 'Assessment of story comprehension deficits after brain damage.', *Journal of Clinical and Experimental Neuropsychology.,* 27(3), pp. 367-384.

Filiatrault-Veilleux, P., Bouchard, C., Trudeau, N. and Desmarais, C. (2015) 'Inferential comprehension of 3-6 year olds within the context of story grammar: a scoping review.', *International Journal of Language and Communication Disorders.,* 50(6), pp. 737-749.

Ford, J. A. and Milosky, L. M. (2003) 'Inferring emotional reactions in social situations: Differences in children with language impairment.' *Journal of Speech, Language and Hearing Research*, 46(1) pp. 21-30.

Freed, J. and Cain, K. (2016) 'Assessing school-aged children's inference-making: the effect of story test format in listening comprehension. *International Journal of Language and Communication Disorders.* DOI: 10.1111/1460-6984.12260

Gardner, M. F. 1993. Test of Auditory Reasoning and processing Skills (TARPS). Northumberland: Ann Arbor Publishers Ltd.

Garson, G. D. (2013) *Vailidity and Reliability. Statistical Associates Publishing*: www.statisticalassociates.com.

Gernsbacher, M. A. and Pripas-Kapit, S. R. (2012) 'Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension.', *Metaphor and Symbol,* 27, pp. 93-105.

Gough, P. and Tunmer, W. (1986) 'Decoding reading and reading disability.', *RASE: Remedial & Special Education.,* 7(1), pp. 6-10.

Gov.uk *Equality Act 2010 legislation.gov.uk*. (Accessed: 20/02 2016).

Gregory, R. J. (2004) *Psychological Testing. History, Principles and Applications (4th Edition).* London: Pearson.

Hadwin, J. A., Howlin, P. and Baron-Cohen, S. (2015) *Teaching children with autism to mind read. The workbook.* Chichester: Wiley-Blackwell.

Happe, F. G. E. (1993) 'Communicative competence and theory of mind in autism. A test of the relevance theory.', *Cognition,* 48, pp. 101-119.

Happe, F., G. (1994) 'An advanced test of theory of mind: understanding of story character's thoughts and feelings by able autistic, mentally handicapped and normal children and adults.', *Journal of Autism and Developmental Disorders,* 24(2), pp. 129-145.

Hatcher, L. (2013) *Advanced Statistics in Research.* Saginaw, MI: Shadow Finch Media.

Hobbs, G. and Vignoles, A. (2007) *Is free school meal status a valid proxy for socio-economic status in schools research? [Online] [Accessed on 22nd July 2016]*.http://webarchive.nationalarchives.gov.uk/20130401151715/http//education.gov.uk/publications/eorderingdownload/cee03-07.pdf

Hoff, E. (2006) 'How social contexts support and shape language development'. *Developmental Review,* 26(1), pp.55-88

Hoffmann, A., Martens, M.A., Fox, R., Rabidoux, P. and Andridge, R. (2013) 'Pragmatic language assessment in William's Syndrome: a comparison of the Test of Pragmatic Language-2 and the Children's Communication Checklist-2. *American Journal of Speech-Language Pathology,* 22(2), pp.198

Howlin, P., Baron-Cohen, S. and Hadwin, J., A. (1998) *Teaching children with autism to mind read: A practical guide for teachers and parents.* Chichester: Wiley.

Huber-Okrainec, J., Blaser, S., E. and Dennis, M. (2005) 'Idiom comprehension deficits in relation to corpus callosum agenesis and hypoplasia in children with spina bifida meningomyelocele.', *Brain and Language.,* 93, pp. 349-368.

Huisingh, R., Bowers, L. and LoGiudice, C. 2006. The Listening Comprehension Test 2. Linguisystems Inc.

IBM 2013. SPSS Statistics. Version 19.: IBM.

Jabrayilov, R., Emons, W.H.M., & Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied Psychological Measurement*, 40, 8, 559-572.

James, S., L. (1986) 'Test Review: Preschool Language Assessment Instrument (PLAI)', *The Reading Teacher,* (December 1986), pp. 344-346.

Johnson, M. 1998. Canterbury and Thanet Verbal Reasoning Assessment. East Kent Community NHS Trust.

Jolliffe, T. and Baron-Cohen, S. (1999) 'The strange stories test: A replication with high-functioning adults with autism or Asperger syndrome.' *Journal of Autism and Developmental Disorders.*, 29(5)

Kaland, N., Moller-Nielsen, A., Smith, L., Mortensen, E. L., Callensen, K. and Gottlieb, D. (2005) 'The strange stories test: a replication study of children and adolescents with Asperger syndrome.', *European Child and Adolescent Psychiatry,* 14, pp. 73-82.

Kana, R. K., Murdaugh, D. L., Wolfe, K. R. and Kumar, S. L. (2012) 'Brain responses mediating idiom comprehension: Gender and hemispheric differences.', *Brain Research,* 1467, pp. 18-26.

Karasinski, C. and Weismer, S. E. (2010) 'Comprehension of inferences in discourse processing by adolescents with and without language impairment.', *Journal of Speech, Language and Hearing Research.,* 53, pp. 1268-1279.

Kerbel, D. and Grunwell, P. (1998) 'A study of idiom comprehension in children with semantic-pragmatic difficulties. Part 1: Task effects on the assessment of idiom comprehension in children.', *International Journal of Language and Communication Disorders.,* 33(1), pp. 1-22.

Kline, P. (2000) *The Handbook of Psychological Testing.* Second Edition. London: Routledge.

Law, J., McBean, K. and Rush, R. (2011) 'Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage.', *International Journal of Language and Communication Disorder,* 46, pp. 657-664.

Lawshe, C. H. (1975) 'A quantitative approach to content validity.', *Personnel Psychology,* 28, pp. 563-575.

LDA (1977) *What Would You Do?* Wisbech: Learning Development Aids.

Lee, S. B., Song, S. H., Ham, J. H., Song, D. H. and Cheon, K.-A. (2015) 'Idiom comprehension deficits in high-functioning autism spectrum disorder using a Korean Autism Social Language Task.', *Yonsei Medical Journal.,* 56(6).

Legler, D., M. (1991) *Don't take it so literally. Reproducible activities for teaching idioms.* Unknown binding.: www.amazon.co.uk.

Leinonen, E. and Letts, C. (1997) 'Why pragmatic impairment? A case study in the comprehension of inferential meaning.', *European Journal of Disorders of Communication.,* 32, pp. 35-51.

Letts, C. and Leinonen, E. (2001) 'Comprehension of inferential meaning in language-impaired and language normal children.', *International Journal of Language and Communication Disorders.,* 36(3), pp. 307-328.

Lovas, G. (2011) 'Gender and patterns of language development in mother-toddler and father-toddler dyads', *First Language,* 31(1), pp. 83- 108.

Lucas, R. and Norbury, C. F. (2015) 'Making inferences from text: It's vocabulary that matters.', *Journal of Speech, Language and Hearing Research.,* 58, pp. 1224-1232.

Lundblom, E. E. G. and Woods, J. J. (2012) 'Working in the classroom: improving idiom comprehension through class wide peer tutoring.' *Communication Disorders Quarterly*, 33(4) pp. 202-219.

Lynch, J. S. and Van den Brock, P. (2007) 'Understanding the glue of narrative structure: Children's on- and off-line inferences about characters' goals.' *Cognitive Development*, 22 pp. 323-340.

Maddocks, R. 2015. Idiom comprehension in typically developing children between the ages of 10:00 and 11:11 years. Unpublished final year dissertation.: Manchester Metropolitan University.

McCall, W. A. (1939) *Measurement.* New York.: Macmillan.

McCartney, R. 2005. High Level Language and Communication Screening Test. Whiston, L35: Knowsley Primary Care Trust.

McCauley, R., J. (2001) *Assessment of Language Disorders in Children.* London: Lawrence Erlbaum Associates.

McDonald, S., Flanagan, S., Rollins, J. *and* Kinch, J., (2003) 'A new clinical tool for assessing social perception after traumatic brain injury'. *Journal of Head Trauma Rehabilitation,* 18, 219–238.

McInnes, A., Humphries, T., Hogg-Johnson, S. and Tannock, R. (2003) 'Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment.', *Journal of abnormal child psychology.,* 31(4), pp. 427-443.

Moran, C. A., Nippold, M. A. and Gillon, M. A. (2006) 'Working memory and proverb comprehension in adolescents with traumatic brain injury: A preliminary investigation.', *Brain Injury.,* 20(4), pp. 417-423.

Murphy, K. R. and Davidshofer, C. O. (2005) *Psychological Testing. Principles and Applications (6th Edition).* New Jersey.: Pearson Education Ltd.

Neale, M. 1997. Neale Analysis of Reading Ability - Revised (NARA-II). Windsor, England.: NferNelson.

Neil, S. J. (2005) 'Research with children: a critical review of the guidelines.', *Journal of Child Health,* 9(1), pp. 46-58.

Newcomer, P. L. and Hammill, D. D. 1997. Test of Language Development - Primary, Third Edition. Austin, TX: Pro-Ed.

NIAR (2010) *Free school meal entitlement as a measure of deprivation. Northern Ireland Assembly Report.*: Northern Ireland Assembly.

Nippold, M., A. (1998) *Later language development: The school-age and adolescent years.* 2nd Edition. Austin, Texas.: PRO-ED inc.

Nippold. Marilyn, A. and Taylor, C., L. (2002) 'Judgments of idiom familiarity and transparency: A comparison of children and adolescents.', *Journal of Speech, Language and Hearing Research.,* 45, pp. 384-391.

Nippold, M. and Duthie, J. K. (2003) 'Mental imagery and idiom comprehension: a comparison of school-age children and adults.', *Journal of Speech, Language and Hearing Research,* 46, pp. 788-799.

Norbury, C. F. and Bishop, D. V. M. (2002) 'Inferential processing and story recall in children with communication problems: a comparison of specific language impairment, pragmatic language impairment and high-functioning autism.', *International Journal of Language and Communication Disorders.,* 37(3), pp. 227-251.

Norbury, C. F. (2004) 'Factors supporting idiom comprehension in children with communication disorders.', *Journal of Speech, Language and Hearing Research.,* 47, pp. 1179-1193.

Norbury, C. and Nation, K. (2011) 'Understanding variability in reading comprehension in adolescents with Autism Spectrum Disorders: Interactions with language status and decoding skill.', *Sceintific Studies of Reading.,* 15(3), pp. 91-121.

Oakhill, J. (1984) 'Inferential and memory skills in children's comprehension of stories.', *British Journal of Educational Psychology.,* 54, pp. 31-39.

Oakhill, J. and Cain, K. (2012) 'The precursors of reading ability in young readers: Evidence from a four-year longitudinal study.', *Scientific Studies of Reading.,* 16(2), pp. 91-121.

O'Hare, A. E., Bremner, L., Nash, M., Happe, F. and Pettigrew, L. M. (2009) 'A Clinical Assessment Tool for Advanced Theory of Mind performance in 5 to 12 Year Olds.', *Journal of Autism and developmental Disorders,* 39(6), pp. 916-928.

Owens, R. E. (2004) *Language Disorders. A functional approach to assessment and intervention.* 4th Edition. London: Pearson Education inc.

Pallant, J. (2013) *SPSS Survival manual.* 5th edn. Berkshire: Open University Press.

Parsons, S. and Branagan, A. (2005) *Language for thinking. A structured approach for young children.* Oxon.: Speechmark Publishing Ltd.

Pearson, R., A. M. and Pillow, B. H. (2016) 'Children and Adults' understanding of faux pas and insults.' *Journal of Educational and Developmental Psychology*, 6(2) pp. 14-28.

Phelps-Terasaki, D. and Phelps-Gunn, T. 2007. Test of Pragmatic Language, 2nd Edition. IL. Linguisystems

Piaget, J. (1965) *The child's conception of the world.* Totowa, NJ.: Littlefield Adams.

Pike, M. M., Barnes, M. A. and Barron, R. W. (2010) 'The role of illustrations in children's inferential comprehension.', *Journal of Experimental Child Psychology.,* 105(3), pp. 243-255.

Pike, M. M., Swank, P., Taylor, H., Landrey, S. and Barnes, M. A. (2013) 'Effect of preschool working memory, language and narrative abilities on inferential comprehension in school-age children with Spina Bifida Myelomeninocele and typically developing children.', *Journal of the International Neuropsychological Society.,* 19(4), pp. 390-399.

Plake, B., S. and Wise, L., L. (2014) 'What is the role and importance of the revised AERA, APA, NCME Standards for Educational and Psychological Testing?', *Educational Measurement: Issues and Practice.,* 33(4), pp. 4-12.

Pring, T. (2005) *Research methods in communication disorders.* London.: Whurr Publishers Ltd.

Puche-Navarro, R. and Millan, R. (2007) 'Inferential functioning in visually impaired children.', *Research in Developmental Disabilities.,* 28, pp. 249-265.

Qualls, C. and Harris, J. (1999) 'Effects of familiarity on idiom comprehension in African American and European American fifth graders.', *Language, Speech and Hearing Services in Schools.,* 30, pp. 141-151.

Qualls, C. D., Lantza, J. M., Pietrzykb, R. M., Blooda, G. W. and Hammera, C. S. (2004) 'Comprehension of idioms in adolescents with language-based learning disabilities compared to their typically developing peers.', *Journal of Communication Disorders.,* 37, pp. 295-311.

Rhea, P. and Norbury, C., F. (2012) *Language disorders from infancy through adolescence: listening, speaking, reading, writing and communicating.* 4th edn. St Louis,Mo.: Elesvier.

Rinaldi, W. 1996. Understanding Ambiguity: An Assessment of Pragmatic Meaning Comprehension.: NFER Nelson.

Rinaldi, W. (2000) 'Pragmatic comprehension in secondary school-aged students with specific developmental language disorder.', *International Journal of Language and Communication Disorders.,* 35(1), pp. 1-29.

Robertson, S.,Kersner, M. and Davies, S. (1995) *From the College of Speech Therapists to the RCSLT. A history of the College 1945-1995.* London. RCSLT.

Ross, E. 2011. Black Sheep Press: Informal assessment of social language and communication skills for children in primary school. 67, Middleton, Cowling, Keighley, W. Yorks, BD22 0DQ, England: Black Sheep Press.

Ryder, N., Leinonen, E. and Schulz, J. (2008) 'Cognitive approach to assessing pragmatic language comprehension in children with specific language impairment.', *International Journal of Language & Communication Disorders.,* 43(4), pp. 427-447.

Saladana, D. and Frith, U. (2007) 'Do readers with autism make bridging inferences from world knowledge?', *Journal of Experimental Child Psychology.,* 96, pp. 310-319.

Schon, D. (1983) *The reflective practitioner. How professionals think in action.* U.S.A.: Basic Books, Inc.

Semel, E., Wiig, E., H and Secord, W. 2006. Clinical Evaluation of Language Fundamentals - Fourth Edition, UK (CELF-4 UK). Oxford, UK: Pearson Education Ltd.

Shields, J. (1991) 'Semantic pragmatic disorder - a right hemisphere syndrome.', *British Journal of Disorders of Communication.,* 26(3), pp. 383-392.

Shields, J., Varley, R., Broks, P. and Simpson, A. (1996) 'Social cognition in developmental language disorders and high-level autism.' *Developmental Medicine and Child Neurology.*, 38 pp. 487-495.

Shore, L.M (2015) 'Talking in class:a study of socio-economic difference in the primary school classroom.' *Literacy*, 49(2), pp. 98-104

Silva, M. and Cain, K. (2015) 'The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension.', *Journal of Educational Psychology.,* 107(2), pp. 321-331.

Simpson, F. 2006. Mount Wilga High Level Language Test - revised edition. Durham, UK: Fiona@conkertree.co.uk.

Spanoudis, G., Natsopoulos, D. and Panayiotou, G. (2007) 'Mental verbs and pragmatic language difficulties.', *International Journal of Language & Communication Disorders.,* 42(4), pp. 487-504.

Speechmark resources: speechmark.net

Spencer, S., Clegg, J., Stackhouse, J. and Rush, R. (2016) 'Contribution of spoken language and socio-economic background to adolescents' educational achievement at age 16 years.', *International Journal of Language & Communication Disorders. doi:10.1111/1460-6984.12264*.

Sperber, D. and Wilson, D. (1997) 'Remarks on relevance theory and the social sciences.', *Multilingua.,* 16, pp. 145-151.

Stansfield, J. and Armstrong, L. (2016) 'A content analysis of the professional journal of the College of Speech Therapists II: Coming of age and growing maturity.', *International Journal of Language & Communication Disorders,* 51(4), pp 478-486.

Stokes, P. (2011) *Key concepts in business and management research methods.* Hampshire: Palgrave Macmillan.

Stokes, P. and Wall, T. (2014) *Research methods. Business Briefings* London: Palgrave.

Super Duper Publications resources: superduperinc.com

Swanell, C. (2014) 'Experts opt for opt-out consent.', *Medical Journal of Australia,* (7).

Tabachnick, B., G. and Fidell, L., S. (2007) *Using multivariate statistics.* 5th edn. London: Pearson.

Tadic, V., Cooper, A., Cumberland, P., Lewando-Hundt, G. and Rahi, J. S. (2013) 'Development of the Functional Vision Questionnaire for children and young people with visual impairment.', *Opthalmolgy.,* 120(12), pp. 2725-2732.

Talarowska, M., Florkowski, A., Orzechowska, A., Zboralski, K., Lechanska, M. D. and Galecki, P. (2012) 'The use of the RHLB battery for the evaluation of the lingual and social skills among psychiatric patients - case study.', *Psychiatr. Pol.,* 46(6), pp. 1089-1098.

Terman, L., M and Merrill, M., A 1960. Stanford-Binet Intelligence Scales: Manual for the 3rd revision. Boston: Houghton Mifflin.

Thomson, A. M., Taylor, R., Fraser, D. and Whittle, I. R. (1997) 'The utility of the Right Hemisphere Language Battery in patients with brain tumours.', *The European Journal of Disorders of Communication.,* 32, pp. 325-332.

Thorum, A. R. 1986. The Fullerton Language Test for Adolescents (Revised). California: California Consulting Psychologists Press Inc.

Trabasso, T., and Magliano, J.P. (1996) 'Conscious understanding during comprehension'. *Discourse Processes*, 21, 255-287.

Van Kleeck, A. (2008) 'Providing preschool foundations for later reading comprehension: the importance of and ideas for targeting inferencing in storybook sharing interventions.', *Psychology in the Schools,* 45(7).

Vasilyeva, M., Waterfall, H. and Huttenlocher, J. (2008) 'Emergence of syntax: commonalities and differences across children.' *Developmental Science,* 11 (1), pp. 84-97

Volden, J and Phillips, L. (2010). 'Measuring pragmatic language in speakers with Autism Spectrum Disorders: Comparing the Children's Communication Checklist -2 and the Test of Pragmatic Language. *American Journal of Speech-Language Pathology,* 19, pp. 204-212

Wechsler, D. 1992. Wechsler Intelligence Scale for Children. Third Edition. Hove.: Psychological Corporation.

WHO ICD10 (2004). *http://apps/who.int/classification/icd10/browse/2016* (Accessed: 20/02 2016).

Whyte, E., M., Nelson, K., E. and Scherf, K. S. (2014) 'Idiom, syntax, and advanced theory of mind abilities in children with autism spectrum disorders.', *Journal of Speech, Language, and Hearing Research.,* 57, pp. 120-130.

Wiig, E., H. and Secord, W. 1989. Test of Language Competence - Expanded Edition.: The Psychological Corporation. Harcourt Brace and Company.

Wiig, E., H. and Secord, W. 1991. Test of Word Knowledge. U.S.A: The Psychological Corporation.

Wimmer, H. and Perner, J. (1983) 'Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception.' *Cognition*, 1(3) pp. 103-128.

Yu, C. H. (2013) *A simple guide to the Item Response Theory and Rasch Modeling*. http://www.creative-wisdom.com (Accessed: 2nd August 2016).

Zachman, L., Jorgensen, C., Huisingh, R. and Barrett, M. 1984. Test of Problem Solving. Ilinois: Linguisystems Inc.

Zachman, L., Barrett, M., Huisingh, R. and Blagden, C. 1991. Test of Problem Solving - Adolescent. East Moline, IL.: LinguiSystems.

Zalla, T., Sav, A. M., Stopin, A., Ahade, S. and Leboyer, M. (2009) 'Faux pas detection and intentional action in Asperger Syndrome. A replication on a French sample.', *Journal of Autism and Developmental Disorders.,* 39(2), pp. 373-82.

Appendix i

DSM-5 (2013) and ICD-10 (2004) Criteria for ASD

DSM-5 (2013)

To receive a diagnosis of ASD a child must show all of the 3 criteria from Section A, at least 2 of the criteria in B, and fulfil C and D.

A. Social communication/interaction deficits.

1. Deficits in social-emotional reciprocity ranging from: abnormal social approach and failure of back and forth conversation; reduced sharing of interests, emotions, and affect and response; total lack of initiation of social interaction.

2. Deficits in nonverbal communicative behaviours used for social interaction ranging from: poorly integrated verbal and non-verbal communication; abnormalities in eye contact and body language; deficits in understanding and use of nonverbal communication; total lack of facial expression or gestures.

3. Deficits in developing and maintaining relationships, appropriate to the child's developmental level (beyond those with caregivers) ranging from: difficulties adjusting behaviour to suit different social contexts; difficulties in sharing imaginative play and in making friends; an apparent absence of interest in people.

B. Restricted, repetitive behaviours and interests (and sensory issues).

1. Stereotyped or repetitive speech, motor movements or use of objects (e.g. motor stereotypies, echolalia, repetitive use of objects or idiosyncratic phrases).

2. Excessive adherence to routines, ritualized patterns of verbal or nonverbal behaviour, or excessive resistance to change (e.g. motor rituals, insistence on same route or food, repetitive questioning or extreme distress at small changes).

3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g. strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).

4. Hyper- or hypo-reactivity to sensory input or unusual interest in sensory aspects of the environment (e.g. apparent indifference to pain/heat/cold, adverse response to specific sounds or textures, excessive smelling or touching of objects, fascination with lights or spinning objects).

C. Symptoms present in early childhood.

D. Symptoms together limit and impair everyday functioning.
The professional who makes the diagnosis must also specify if the ASD: occurs with or without accompanying intellectual or language impairment or catatonia; is associated with a known medical or genetic condition or environmental factor; or is associated with another neurodevelopmental, mental or behavioural disorder

<u>ICD-10 (2004)</u>

The ICD 10 defines childhood autism in section F84.0 as:
A. The presence of abnormal or impaired development before the age of 3 years, in at least one of the following areas:
1. Receptive or expressive language as used in social communication.
2. The development of selective social attachments or of reciprocal social attachment.
3. Functional or symbolic play.

B. Qualitative abnormalities in reciprocal social interaction, manifest in at least one of the following areas:
1. Failure to adequately use eye-to-eye gaze, facial expression, body posture and gesture to regulate social interaction.
2. Failure to develop (in a manner appropriate to mental age, and despite ample opportunities) peer relationships that involve a mutual sharing of interests, activities and emotions.
3. A lack of socio-emotional reciprocity as shown by an impaired or deviant response to other people's emotions; or lack of modulation of behaviour according to social context, or a weak integration of social, emotional and communicative behaviours.

C. Qualitative abnormalities in communication manifest in a least two of the following areas:
1. A delay in, or total lack of development of spoken language that is not accompanied by an attempt to compensate through the use of gesture or mime as alternative modes of communication (often preceded by a lack of communicative babbling).

2. Relative failure to initiate or sustain conversational interchange (at whatever level of language skills are present) in which there is a reciprocal to and fro responsiveness to the communication of the other person.
3. Stereotyped and repetitive use of language and idiosyncratic use of words or phrases.
4. Abnormalities in pitch, stress, rate, rhythm and intonation of speech.
D. Restricted, repetitive, and stereotyped patterns of behaviour, interests and activities, manifest in at least two of the following areas:
1. An encompassing preoccupation with one or more stereotyped and restricted patterns of interest that are abnormal in content or focus; or one or more interests that are abnormal in their intensity and circumscribed nature although not abnormal in their content of focus.
2. Apparently compulsive adherence to specific, non-functional, routines or rituals.
3. Stereotyped and repetitive motor mannerisms that involve either hand or finger flapping or twisting, or complex whole body movements;
4. Preoccupations with part-objects or non-functional elements of play materials (such as their odour, the feel of their surface, or the noise or vibration that they generate);
5. Distress over changes in small, non-functional, details of the environment.
E. The clinical picture is not attributable to other varieties of pervasive developmental disorder; specific developmental disorder of receptive language (F80.2) with secondary socio-emotional problems; reactive attachment disorder (F94.1) or disinhibited attachment disorder (F94.2); mental retardation (F70-F72) with some associated emotional or behavioural disorder; schizophrenia (F20) of unusually early onset; and Rett's syndrome (F84.2).
The latest version of the ICD (ICD11) was due to be available in 2015 but has been delayed until 2017. The revision is likely to give four categories under the ASD umbrella:

1. Autism (Prototypical ASD) defined as an impaired capacity for reciprocal socio-communicative interaction, together with restricted interests and repetitive behaviours.
2. Disintegrative disorder (where the child loses previously acquired developmental skills up to 3 years).

3. Social reciprocity disorder (formerly Asperger syndrome).
4. Rett syndrome.

APA (2013) *DSM5* www.dsm5.org (Accessed: 20/02 2016).

WHO ICD10 (2004). *http://apps/who.int/classification/icd10/browse/2016* (Accessed: 20/02 2016).

Appendix ii

Further examples of Happe's (1994) Physical and Mental State Strange Stories

Physical state story examples

1) The architect Ken Peterson is known as a person rich in ideas. He works with Solnes, a master builder who has his office in the town. He goes to Solnes almost daily with new ideas about how to build bigger and better buildings.
The idea rich architect uses steel and glass as construction materials, because they are the materials that can give the most protection against storms and bad weather. With these materials it is possible to build fine, big buildings. Wooden material and roof tiles are well suited for the construction of normal single-floored dwellings, he says.

Question: Why doesn't architect Peterson use wooden material and roof tiles when projecting high buildings?

2) John Frank is 25 years old and tunes all kinds of pianos. He is the only piano tuner in the town, and his clients must normally wait a week before he can do the work they desire. Usually he carries out his work in the homes of the clients.
One day an old lady calls him. She presents herself as Mrs.Agnes Lind, and says that she would like to have her piano tuned. She tells John her address. Some minutes previously a small job John was to do this morning in the concert house had been cancelled. A few minutes later he arrives at Mrs. Lind's villa.

Question: Why doesn't Mrs. Lind have to wait a week to have her piano tuned?

3) David Swenson is broke at the moment, because he has just paid some large bills. One day after filling his old, but well-maintained car with petrol he falls for the temptation of driving off without paying. The attendant at the station is busy with another customer, at the same time his telephone is ringing and a mechanic in the garage is calling him.

Question: Why doesn't the serving attendant at the station stop Swenson when he drives away after not having paid for his petrol?

4) One late, dark autumn evening the 14 year old Paul is going along some scary town streets with his mother. They have heard and read of people who have been robbed of their money in this area.
Earlier that day Paul's mother has been to the bank and made a withdrawal of 8000 NOK. She has placed the money in an inside coat pocket instead of in her handbag. Her old washing machine broke down for good a couple of days ago, and she has to buy a new one in one of the coming days.

Question: Why has Paul's mother been to the bank and made a withdrawal?

5) Two enemy powers have been at war for a very long time. Each army has won several battles, but now the outcome could go either way. The forces are equally matched. However, the Blue army is stronger than the Yellow army in foot soldiers and artillery. But the Yellow army is stronger than the Blue army in air power. On the day of the final battle, which will decide the outcome of the war, there is a heavy fog over the mountains where the fighting is about to occur. Low-lying clouds hang above the soldiers. By the end of the day the Blue army have won.

Question: Why did the Blue army win?

6) Sarah is very long-sighted. She has only one pair of glasses, which she keeps losing. Today she has lost her glasses again and she needs to find them. She had them yesterday evening when she looked up the television programmes. She must have left them somewhere that she has been today. She asks Ted to find her glasses. She tells him that today she went to her regular early morning keep fit class, then to the post office, and las to the flower shop. Ted goes straight to the post office.

Question: Why is the post office the most likely place to look?

<u>Mental State Story examples</u>

<u>Lie (Dentist)</u>

John hates going to the dentist because every time he goes to the dentist he needs a filling, and that hurts a lot. But John knows that when he has toothache, his mother always takes him to the dentist. Now John has bad toothache at the moment, but when his mother notices he is looking ill and asks him ‘‘Do you have toothache, John?’’. John says ‘‘No, Mummy’’.

1. Is it true what John says to his mother?

2. Why does John say this?

<u>White Lie (Hat)</u>

One day Aunt Jane came to visit Peter. Now Peter loves his aunt very much, but today she is wearing a new hat; a new hat which Peter thinks is very ugly indeed. Peter thinks his aunt looks silly in it, and much nicer in her old hat. But when Aunt Jane asks Peter, ‘‘How do you like my new hat?’’ Peter says, ‘‘Oh, it’s very nice’’.

1. Was it true what Peter said?

2. Why did he say it?

<u>Misunderstanding (Glove)</u>

A burglar who has just robbed a shop is making his getaway. As he is running home, a policeman on his beat sees him drop his glove. He doesn’t know the man is a burglar, he just wants to tell him he dropped his glove. But when the policeman shouts out to the burglar, ‘‘Hey you, Stop!’’, the burglar turns round, sees the policeman and gives himself up. He puts his hands up and admits that he did the break-in at the local shop.

1. Was the policeman surprised by what the burglar did?

2. Why did the burglar do this, when the policeman just

wanted to give him back his glove?

<u>Persuasion (Kittens)</u>

Jill wanted to buy a kitten, so she went to see Mrs. Smith who had lots of kittens she didn't want. Now Mrs. Smith loved the kittens and she wouldn't do anything to harm them, though she couldn't keep them all herself. When Jill visited she wasn't sure she wanted one of Mrs. Smith's kittens, since they were all males and she had wanted a female. But Mrs. Smith said, "If no one buys the kittens, I'll just have to drown them!''

1. Was it true what Mrs. Smith said?

2. Why did Mrs. Smith say this to Jill?

<u>Contrary Emotions (Swings)</u>

Today, Katy wants to go on the swings in the playground. But to get to the playground she knows she has to pass old Mr. Jones house. Mr. Jones has a nasty fierce dog and every time Katy walks past the house, the dog jumps up at the gate and barks. It scares Katy awfully and she hates walking past the house because of the nasty dog. But Katy does so want to play on the swings. Katy's mother asks her ''Do you want to go out to the playground?'' Katy says ''No''.

1. Is it true what Katy says?

2. Why does she say she doesn't want to go to the playground, when she so wants to go on the swings that are there?

<u>Joke (Haircut)</u>

Daniel and Ian see Mrs. Thompson coming out of the hairdressers 1 day. She looks a bit funny because the hairdresser has cut her hair much too short. Daniel says to Ian, ''She must have been in a fight with a lawnmower!’’

1. Is it true what Daniel says?

2. Why does he say this?

Figure of Speech (Cough)

Emma has a cough. All through lunch she coughs and coughs and coughs. Father says ‘‘Poor Emma, you must have a frog in your throat!’’

1. Is it true what Father says to Emma?

2. Why does he say that?

Double Bluff (Ping-Pong Bat)

Simon is a big liar. Simon’s brother Jim knows this, he knows that Simon never tells the truth! Now yesterday Simon stole Jim’s ping-pong bat and Jim knows Simon has hidden it somewhere, though he can’t find it. He’s very cross. So he finds Simon and he says ‘‘Where is my pingpong bat? You must have hidden it either in the cupboard or under your bed, because I’ve looked everywhere else. Where is it, in the cupboard or under your bed?’’ Simon tells him the bat is under his bed.

1. Was it true what Simon told Jim?

2. Where will Jim look for his ping-pong bat?

3. Why will Jim look there for his bat?

Appearance/Reality (Santa Claus)

On Christmas Eve, Alice’s mother takes her to the big Department store in town. They go to look in the toy department. In the toy department Mr. Brown, Alice’s next door neighbour, is dressed up as Santa Claus, giving out sweets to all the children. Alice thinks she recognises Mr. Brown, so she runs up to him and asks ‘‘Who are you?’’ Mr. Brown answers ‘‘I’m Santa Claus!’’.

1. Is it true what Mr. Brown says?

2. Why does he say this?

<u>Forget (Doll)</u>

Yvonne is playing in the garden with her doll. She leaves her doll in the garden when her mother calls her in for lunch. While they are having lunch, it starts to rain. Yvonne's mother asks Yvonne ''Did you leave your doll in the garden?'' Yvonne says ''No, I brought her in with me, Mummy''.

1. Is it true what Yvonne says?

2. Why does Yvonne say this?

Happe, F., G. (1994) 'An advanced test of theory of mind: understanding of story character's thoughts and feelings by able autistic, mentally handicapped and normal children and adults.', *Journal of Autism and Developmental Disorders,* 24(2), pp. 129-145.

Appendix iii

Further details of current inferential and idiom comprehension assessments

Inferential comprehension

Liverpool Language Screen (Unpublished data)

A few questions had supporting pictures, given in brackets below.

1.Taking Another's Perspective

1. If you hurt your friend would s/he be happy?
2. If you were cold, would you take your gloves off?
3. If you were hungry would you finish your dinner?
4. If you were a fish would you be able to fly?

2.Predictions

5. What would happen to a flower if it didn't get any water?
6. What happens to a window if a cricket ball hits it?
7. What might happen if you let a baby get hold of a knife?
8. What would happen to an egg if you dropped it on the floor?

3.Obstacles to Action and Experience

9. Why can't he put the shoe on? (Picture of a boy trying to put a baby's shoe on).
10. Why isn't the telly working? (Picture of a television that is not plugged in).
11. Why won't the scissors cut? (Picture of scissors with half a blade missing).
12. Why can't he see who he's caught? (Picture of a blindfolded boy catching hold of a girl's pony tail).
13. Why can't you read in the dark?
14. Why can't you make a snowman in the summer time?
15. Why is it silly to go shopping without any money?
16. Why is it silly to hang washing out to dry when it's raining?

4.Cause – Effect/ Prevention

17. Why mustn't you play with matches?
18. Why do you wear an apron/shirt when you are cooking/painting?
19. Why do people have umbrellas?
20. Why do people cry?
21. Why do we need a towel when we go swimming?

22. Why does a car need brakes?
23. Why is this girl shouting ‘Help’? (picture of a girl who has fallen through a hole in the ice).
24. Why is this lady pushing the car? (picture of a lady pushing a broken down car).

5.Formulating solutions to a Problem

25. How can you talk to somebody who is hundreds of miles away?
26. How would you get to the top floor of a very tall block of flats?
27. How do we keep milk cold?
28. How could you find out what your teacher looked like when s/he was a little girl/boy?

6.Explanation of an Inference

29. How can you tell that the children have been eating biscuits? (Picture of 2 children with an open packet of biscuits and a plate each with crumbs on).
30. How can you tell that it’s raining outside? (Picture of 2 dripping wet children standing in the hallway).
31. How can you tell that the music is too loud? (Picture of a loud speaker blaring out music and a boy grimacing and covering his ears with his hands).
32. How can you tell that the girl doesn’t like this pie? (Picture of a girl sticking out her tongue and pushing a pie away from herself).
33. How can you tell that someone is tired?
34. How can you tell that somebody is a policeman?
35. How can you tell what’s inside a packet at the supermarket?
36. How could you tell that a dog wasn’t hungry?

7.Construction of Objects

37. Why are stamps sticky on the back?
38. Why are windows made out of glass and not out of bricks?
39. Why do birds need wings?
40. Why do people need teeth?

8.Story Comprehension

‘A great big lion caught a little tiny mouse in its paw and was going to eat it. “Please don’t kill me mighty lion,” said the little mouse. “I am too small to be a proper dinner for you.”

"Alright," said the lion. "I'll let you go."

"Thank you, thank you," squeaked the mouse. "One day I will do something to help you."

"How ever could you help me?" asked the lion, roaring with laughter.

But not long afterwards some hunters caught the lion and tied it up in a net. The mouse came creeping up and gnawed a big hole in the net with its sharp teeth, so that the lion could escape.

41. Did the mouse do something to help the lion?
42. Did the lion think that the mouse would be able to help it?
43. When the lion caught the mouse, what was the mouse afraid would happen?
44. Why did the lion let the mouse go?
45. Who caught the lion?
46. What did they do to him?
47. What do you think they would have done to him next?
48. Why did the mouse want to help the lion?
49. How did the mouse make a hole in the net?

All correctly answered questions scored 1 point except 13, 14, 15, 20, 25, 33, 37, 40, 44, 45, 46, 48 which could score 1 or 2. Marking criteria were agreed upon. The maximum total score was 61.

No major gender differences were found in the scores so the boys and girls scores were combined.

Sub-test and Total Test Mean Scores for each age group

Age group	Sec 1 /4	Sec 2 /4	Sec 3 /11	Sec 4 /9	Sec 5 /5	Sec 6 /9	Sec 7 /6	Sec 8 /13	Total Test /61
4:00-4:05	4	4	8	7	2	4	3	6	39
4.06-4:11	4	4	8	7	2	6	4	8	42
5:00-5:05	4	4	9	7	2	6	4	7	44
5:06-5:11	4	4	10	8	3	7	4	10	50
6:00-6:05	4	4	10	8	4	8	5	10	51
6:06-6:11	4	4	10	8	4	8	5	10	53

Language for Thinking. A structured approach for young children (Parsons and Branagan, 2005)

Assessment i) Cinema

Text: On Saturday John went to the cinema. John was excited. John went with his mum. John's mum bought the tickets. They went and found their seats. There were lots of other people there too. Everyone laughed at the film about aliens.

Questions

1. Tell me what happened in this story.
2. Where are John and his Mum?
3. Who went with his Mum?
4. Who is in the film?
5. When did John go to the cinema?
6. Finish this: 'John felt very .....'
7. Do they like the film? (not scored). Why do they like it?
8. How do you know they like it?
9. What are the children looking at?
10. What's the difference between an alien (clown/puppet) and a dog?
11. What is a cinema (circus/puppet show)?

12. Which is better, TV or cinema (circus/ puppet show)? (not scored). Why?
13. What is John sitting on?
14. Why shouldn't John stand up when the film is on?
15. John is hungry. What could he say?
16. John doesn't have enough money to buy popcorn. What could he do?
17. Tell me how you buy a ticket for the cinema?
18. What will happen after the film has finished?

Total maximum score is 54 (18 times 3).

Interpretation of total score

Score of 0-9: do not start the programme. Try early developmental teaching methods and/or consult a SLT.

Score of 10-23: Start at Language Level A

Score of 24-35: Start at Language Level B

Score of 36-47: Start at Language Level C

Score of 48+ Module completed. Proceed to next module.

The other 2 assessments for monitoring progress contain the same questions in slightly different contexts (ii) At the circus and iii) Puppet show).

Tests not based on the Language of Learning model

Standardised Assessments

The Listening Test (Barrett et al., 1992)

C: Concepts

This is assessed by the student pointing to the appropriate part of a picture.

The concepts assessed are:

- Something you find in a classroom
- Middle
- Healthy
- Neither....nor
- All the letters
- On the left

- Cheapest ball
- Following directions on a map (at corner, go north, turn left, go halfway down, in front of you)
- Before/ first
- The day after tomorrow
- Odd and even-numbered
- A dozen
- Two weeks ago/ Older
- Same number of wheels as...
- The 3 holidays

<u>E: Story Comprehension (3 questions asked per story)</u>

<u>First story</u>

Once upon a time, a young princess was walking through a forest in her kingdom. She sat down on a tree stump to rest. A wise old owl flew out of the forest and landed near the princess. The owl said, "Princess, I will give you three wishes for anything you would like."

1. How do you know the story is make believe?

A. Owls can't talk/ wishes can't come true.

First, the princess wished for enough food to feed everyone in the kingdom. Her second wish was for new shoes for all the children. Her last wish was for trees that grew bubble gum.

2. How do you know the princess wasn't selfish?

A. Wishing for things for other people.

The owl granted all three wishes in an instant. He was so pleased with the princess' wishes that he gave her something extra – a beautiful magic ring. He said "Whenever you need anything, just rub this magic ring, and I'll come to help you." Everyone lived happily ever after.

3. What's a good title for the story?

A. Magic owl/ magic ring/ princess/ wishes

<u>The Listening Comprehension Test 2 (Huisingh et al., 2006).</u>

<u>Passage 3</u>

The first year in a new land was hard for the Pilgrims. Many of them got sick. Their cell phones would not work. They had very little food. It was a tough life, but they were glad to be free.

A. Main Idea

9. What am I talking about? *Only acceptable responses*: any reference to hardships, hard/tough/rough life, Pilgrims

B. Details

10. Why was the first year a hard year for the Pilgrims? *Only acceptable responses*: any reference to they got sick/had little food/supplies

C. Reasoning

11. What doesn't make sense about this story?

Only acceptable responses: any reference to cell phones

D. Vocabulary

12. What is another word for *tough* in this sentence? *It was a tough life.*

Only acceptable responses: hard, harsh, rough, challenging, dangerous, difficult, not easy, complicated, gruelling

E. Understanding Messages is the last subtest on the assessment. It requires the child to interpret "real-life" messages such as public announcements. An example is given below:

Message:

Parent conferences are next week, so school will dismiss 30 minutes early on Tuesday, Wednesday, and Thursday.

Question: What days are parent conferences?

Question: Why will students be dismissed early next week?

Mount Wilga High Level Language Test (Christie et al., 1986) and the revised version (Simpson, 2006)

There are 8 main test sections with sub-sections within them. These are:

1. Naming skills covering four types of naming: A. Naming from description (e.g. name an object which protects you from the rain); B. Category naming (e.g. if 'table' belongs to the category of 'furniture', what group does the following belong to: e.g. Jupiter); C. Association naming (e.g. in 1 minute name as many animals as you can think of); D. Divergent semantics (e.g. give me 2 different examples of things that fizz).

2. Verbal explanation covering: E. Definitions (Explain what the following words mean: What is a: clock); F. Differentiation (What is the main difference between these things? -Telephone/letter); G. Absurdities

H. Idioms; I. Verbal reasoning.

3. Planning covering: J. Jumbled sentences (e.g. rearrange the sentences to make a correct sentence – I sisters two brother and one have); K. Sentence construction (e.g. include the following words into a sentence – Crop although drought); L. Sequencing (e.g. tell me how to make a cup of coffee).

4. Auditory memory covering: M. Sentence repetition (e.g.'They went on a picnic'); N. Passage recall.

5. Auditory comprehension covering: O. Auditory Comprehension of a heard passage (Ten questions mostly factual but the final two questions involve inferential comprehension); P. Logico-semantic relationships (e.g. If 'car' is the whole, 'wheels' will be part of it: so if 'hand' is the whole, what are the parts?); Q. Logico-grammatical relationships (e.g. Is someone's sister their father's daughter or their daughter's father?; David is taller than Michael. Who is the shorter, etc.).

6. Reading comprehension covering: R. Factual paragraph (understanding a silently read newspaper article); S. Inferential Paragraphs.

7. Written expression covering: T. Dictation (write the sentences that I read to you); U. Composition (Write a paragraph explaining how you would set about organising a holiday).

8. Numeracy covering: V. Problem solving (mostly maths problems).

Nonstandardised Assessments

Canterbury and Thanet Verbal Reasoning Assessment (Johnson, 1998)

Age levels by which typically developing children should be able to answer the questions are given for the various question types. Some examples of the sequential logic questions are given below:

By 5 years

Inference (completes a story by supplying 5 sensible words out of a possible 10).

By 6 years

Cause-effect:

"Today I heard a man say to a little boy when he was going out, "It's raining hard Johnny, so...............". Then I could not hear the rest.

What do you think he said next?

Deductions:

Wagwums: Present pictures of a funny creature. "All wagwums have only one leg and no noses"

Is this a wagwum ? (the creature has 2 legs and a nose)

By 7 years

Inference:

It looks like rain, but I shall stay indoors today. Shall I want an umbrella?

How do you know? Can you explain why?

By 8 years

Verbal absurdities:

"I'm going to tell you something and then ask you what was silly about it"

a)Some people found a man locked in his room with his hands tied behind him and his feet tied together. They think he locked himself in.

b)A wheel came off Mr Smith's car. As he could not get the wheel back on by himself, he drove his car to the garage to get it mended. (If a child begins to talk about a tyre, stress that it was a WHEEL and repeat the sentence).

By 8-9 years

Cause-Effect

- What would you do if you were in the street and found a baby that was lost?

By 9 years

Verbal Absurdities:

a) I saw a smart man who was walking down the road with his hands in his pockets and twirling an umbrella.
b) A father wrote to his son 'Here is fifty pounds. If you do not receive this letter, please let me know straightaway.'

By 10 years

Cause-effect:

Give 2 reasons why most people would rather have a car than a bicycle.

Verbal absurdities:

c) They began the meeting late, but they set the hands of the clock back so that the meeting would finish before sunset.

By 11 years

Verbal Absurdities:

d)A man was taking a heavy bag of grain to town on his horse. He wanted to make the load easier for the horse, so he sat on his horse and lifted the bag onto his own shoulder,

By 12 - 13years

Verbal absurdities:

e) One day we saw several icebergs that had been entirely melted by the warmth of the Gulf Stream.

By 14 years

Deduction:

"My house was burgled last Saturday. I was at home all of the morning but out during the afternoon until 5 o'clock. My father left the house at 3 o'clock and my brother was there until 4 o'clock. At what time did the burglary take place?

Idioms/ Non-literal understanding

Standardised assessments

Test of Language Competence (Wiig and Secord, 1989)

The other idioms assessed at level 1 are:

go fly a kite; I got the jump on X; she has a way with dogs; I fell behind; came apart at the seams; you have to be on your toes now; X saw red; they are cut from the same cloth; we are going in circles; you are always making waves; took the wind out of X's sails; we cut corners; he marches to a different drummer.

.

The other idioms assessed at level 2 are:
rough sailing ahead; holding all the aces; zero in on; high on the totem pole; stew over something; I can't swallow that; new blood; up in the air; easily crushed; X is transparent; it's all behind us

Test of Word Knowledge (Wiig and Secord, 1991)

The other idioms covered in this sub-test are: litterbug; all ears; work your fingers to the bone; it goes in one ear and out the other; acting up; going back on your word; getting even; giving a pat on the back; go along with; shake a leg; bite off more than you can chew; bear with; out of place; like a ball of fire; have a way with words; call the shots; fish for something; hold off; grin and bear it; let the cat out of the bag; face to face; cost an arm and a leg; car pool; spring chicken; don't cross the bridge before you come to it; give someone the cold shoulder; close at hand; off the top of your head; nip it in the bud; in the dark; beat your head against a wall; on the way up; easily crushed; come to terms; let bygones be bygones; die on the vine; set one's sights; get the jump on; wise up; make good; wet behind the ears; get the ball rolling.

Appendix iv

Application Number…………….. (**faculty coding**)

(Nov 2006)

Date………………………………..

Manchester Metropolitan University

MANCHESTER METROPOLITAN UNIVERSITY FACULTY OF HEALTH, PSYCHOLOGY AND SOCIAL CARE

APPLICATION FOR ETHICAL APPROVAL

Introduction

All university activity must be reviewed for ethical approval. In particular, all undergraduate, postgraduate and staff research work, projects and taught programmes must obtain approval from their Faculty Ethics committee (or delegated Departmental Ethics Committee).

APPLICATION PROCEDURE

The form should be completed legibly (preferably typed) and, so far as possible, in a way which would enable a layperson to understand the aims and methods of the research. Every relevant section should be completed. Applicants should also include a copy of any proposed advert, information sheet, consent form and, if relevant, any questionnaire being used. The Principal Investigator should sign the application form. Supporting documents, together with one copy of the full protocol should be sent to the Administrator of the appropriate Faculty Ethics Committee.

Your application will require external ethical approval by an NHS Research Ethics Committee if your research involves staff, patients or premises of the NHS (see guidance notes)

Work with children and vulnerable adults

You will be required to have a Criminal Disclosure, if your work involves children or vulnerable adults.

The Faculty Academic Ethics Committee is expected to meet once or twice a term and will respond as soon as possible. Applications that require approval

by an NHS Research Ethics Committee or a Criminal Disclosure will take longer - perhaps 3 months.

1. DETAILS OF APPLICANT (S)
1.1 Principal Investigator: (Member of staff responsible for work)
Name, qualifications, post held, tel. no, e-mail
Anne Hewitt, Senior Lecturer in Speech Pathology and Therapy, MMU.
BSc honours in Speech Sciences. M.Ed in Special Needs and Dyslexia.

1.2 Co-Workers and their role in the project. (e.g. students)
Details (Name, tel. no, email, Course)
None

1.3 University Department/Research Institute

MMU, Professional Registration Division, Speech Pathology and Therapy Department.

2. DETAILS OF THE PROJECT
2.1 Title:
"The development of a British assessment of inferential comprehension for 7-11 year-old typically-developing children."

2.2 Description of Project: (please **outline** the scientific background and the purpose of the research project, 250 words max.). If applicable, please state the hypothesis of your study. Otherwise clearly state its aim.
The aims of the study are:

- **To develop a British assessment of inferential comprehension for 7-11 year-old children.**
- **To provide initial standardisation, reliability and validity data on this assessment.**

Describe what type of study this is (e.g. qualitative or quantitative; also indicate how the data will be analysed) Additional sheets may be attached.
This is a quantitative study.
<u>Methods</u>

- Inferential comprehension questions will be drawn from the literature, expanded upon and made into a pilot assessment.
- The assessment will be a combination of questions based on picture interpretation; questions from verbal scenarios where making inferences from information not explicitly given is necessary. It may also include a social

cognition sub-section (with questions requiring the children to take someone else's perspective).

- The pilot assessment will be trialled on 70 boys and 70 girls (10 per age band) aged 5:00-11:11 (younger children are included to test for floor effects).
- The best performing items from the trial version will be selected for the final test version using statistical analysis, e.g. Cronbach's Alpha coefficient. Detailed scoring guidelines will be drawn up.
- A stratified sample of mainstream primary schools in the North West will be approached to participate in the study. Local Authority ethical guidelines will be followed.
- A random sample of 500 children whose language is developing normally (50 boys and 50 girls for each age group) will be made. The 500 children will be assessed on the final test version.
- Normative scores will be derived from the data (raw score means and standard deviations by age group). Raw scores will be converted to standard scores for each age group.
- Inter- and intra-rater reliability will be assessed by objective assessors examining a sample of video-recorded assessments across the age ranges.
- Concurrent validity will be assessed by comparing the scores of the Inferential Comprehension Test with the scores on the BPVS II (Dunn and Whetton, 1997) and the Inferential Comprehension sub-test of the ACE (Adams et al 2001) using Peasrson's r.
- The internal consistency of the test will be assessed by Cronbach's Alpha coefficient.
- Test-retest reliability may be measured depending on what the literature reveals about the rate of change of inferential comprehension.

2.3 Are You Going To Use A Questionnaire? **NO**

- Please attach a copy if you consider it will raise ethical issues

2.4 Start Date / Duration Of Project.

Data collection to start in 2010 (ideally January but could be later).

2.5 Location Of Where The Project And Data Collection Will Take Place.

Mainstream primary schools in Merseyside and Manchester.

2.6 Nature/Source of Funding.

MMU is funding the PhD fees. No other funding is available.

2.7 Are There Any Regulatory Requirements? NO
If yes, please give details, eg. from relevant professional bodies

3. DETAILS OF PARTICIPANTS

3.1/3.2/3.3/ How many/ Age/ Sex

> **The pilot assessment will be trialled on 70 boys and 70 girls (10 per age band) aged 5:00-11:11.**
>
> **A random sample of 500 children whose language is developing normally (50 boys and 50 girls for each age group 7-11 years) will be made. The 500 children will be assessed on the final test version.**

3.4 How will they be recruited?
(Attach a copy of any proposed advertisement)

A stratified sample of mainstream primary schools in the North West will be approached to participate in the study. Local Authority ethical guidelines will be followed.

3.5 Status of participants (e.g. students, public, colleagues, children, hospital patients, prisoners, including young offenders, participants with mental illness or learning difficulties.)

School children aged 5-11 years.

3.6 Inclusion and exclusions from the project (indicate the criteria to be applied).

Only children whose language is developing typically will be included.

3.7 Payment to volunteers (indicate any sums to be paid to volunteers).

No payment will be made.

3.8 **STUDY INFORMATION**:
Have you provided a study information sheet for the participants? **YES**
Please attach a copy of this information sheet

3.9 **CONSENT**:
(A written consent form for the study participants MUST be provided in all cases, unless the research is a questionnaire.)

Have you produced a written consent form for the participants to sign for your records? **YES**

Please attach your consent form.

4. RISKS AND HAZARDS

4.1 What are the risks to the participants? (Give details of the procedures and processes to be undertaken.)

None.

4.2 State precautions to minimise the risks and possible adverse events.

N/A

4.3 What discomfort (physical or psychological) danger or interference with normal activities might be suffered by the participant(s)?

A child may be anxious about being assessed.

4.4 State precautions been taken to minimise them:

If any child exhibits or expresses any anxiety about being assessed he/she will not have to come for or complete the assessment.

5. WHAT ETHICAL ISSUES DO YOU THINK YOUR STUDY WILL RAISE?

None apart from the one mentioned in 4.3/ 4.4

6. SAFEGUARDS /PROCEDURAL COMPLIANCE

6.1 Confidentiality

(a) Indicate what steps will be taken to safeguard the confidentiality of participant's records. If the data is to be computerised, it will be necessary to ensure compliance with the requirements of the Data Protection Act.

All information about the children will be handled in confidence.
Any information about the children which leaves the school will have his/her name, address and date of birth removed so that he/she cannot be identified.
The assessment forms and all other data from the study will be kept in a locked filing cabinet and the information from these will only be available to the researcher and her supervisory team.

(b) If you are intending to make any kind of audio or visual recordings of the participants, please answer the following questions:

a. How long will the recordings be retained and how will they be stored?

Audio recordings and some video/DVD recordings will be made of the assessments. They will be kept until all the data has been collected and analysed (by 2014 at the latest).

b. How will they be destroyed at the end of the project?

The recordings will be broken into pieces and put in a confidential bin.

c. What further use do you intend to make of the recordings?

None.

6.2 INSURANCE

Are there any insurance or indemnity arrangements in place in the case of negligent or non-negligent harm, other than normal University policies?

NO

Please note: the University holds insurance policies that will cover claims for negligence arising from the conduct of the University's normal business, which includes research carried out by staff and by undergraduate and postgraduate students as part of their course. This does **not** extend to clinical negligence..

6.3 NOTIFICATION OF ADVERSE EVENTS

(Indicate precautions taken against adverse reactions.)

If adverse events do occur, please state the processes/procedures in place to respond to these.

In the case of clinical research, you will need to abide by specific guidance. This may include notification to GP and ethics committee. **Please seek guidance for up to date advice**, eg. see the COREC website at www.corec.org.uk

I will follow the guidelines given in the above website and will seek support from my supervisors.

SIGNATURE OF PRINCIPAL INVESTIGATOR/
PROGRAMME LEADER (for taught programmes) : DATE:

....
....

SIGNATURE OF ETHICS COMMITTEE CHAIRPERSON: DATE:

....

APPENDIX

Checklist of attachments needed:

1. Participant's consent form
2. Participant's information sheet
3. Full protocol
4. Advertising details
5. Insurance notification forms

Example of a model consent form

Title of project

X has explained the nature of the research project and what I would need to do as a volunteer. I have received a copy of the study information sheets, which I have had explained to me. I acknowledge the risks associated with the study and they have been explained to me. Having had y period to consider my decision I am happy to consent to take part in the study and I understand I am free to withdraw at any time. I understand I will receive any payments offered to me up to the time of withdrawal.

I confirm that the facts needed for recruitment are accurate and I will observe all safety precautions listed in the information sheet.

Signed: …………………………………………………Date………..

NAME: ……………………………………………………………

Witnessed: ……………………………………………...…Date……….

NAME: ……………………………………………………………

Appendix v

ETHICS CHECK FORM

This checklist must be completed for every project. It is used to identify whether there are any ethical issues associated with your project and if a full application for ethics approval is required. If a full application is required, you will need to complete the 'Application for Ethical Approval' form and submit it to the relevant Faculty Ethics Committee, or, if your research falls within the NHS and social care, you will need to obtain the required application form from the National Research Ethics Service available at www.nres.npsa.nhs.uk/ and submit it to a local NHS REC.

Before completing this form, please refer to the University's Academic Ethical Framework (www.rdu.mmu.ac.uk/ethics/mmuframework) **and the University's Guidelines on Good Research Practice** (www.rdu.mmu.ac.uk/rdegrees/goodpractice.doc)**.**

Project and Applicant Details

Name of applicant:	Anne Hewitt
Telephone Number:	
Email address:	
Status: **(please circle as appropriate)**	Undergraduate Student Postgraduate Student **Staff**
Department:	Professional Registration Division
Programme of study:	Speech Pathology and Therapy
Name of supervisor:	Jois Stansfield Juliet Goldbart
Project Title:	The development of a British assessment of inferential comprehension for 7-11 year-old typically-developing children.
Does the project require NHS Trust approval? If yes, has approval been granted by the Trust? Attach copy of letter of approval.	NO

Ethics Checklist

Please answer each question by ticking the appropriate box:

	Yes	No	N/A
1. Will the study involve recruitment of patients or staff through the NHS, or involve NHS resources? If yes, you may need full ethical approval from an NHS.		√	
2. Does the study involve participants who are particularly vulnerable or unable to give informed consent (e.g. children, people with learning disabilities, your own students)? **YES**			
3. Will the study require the co-operation of a gatekeeper for initial access to the groups or individuals to be recruited (e.g. students at school, members of self-help group, nursing home residents)?	√		
4. Will the study involve the use of participants' images or sensitive data (e.g. participants personal details stored electronically, image capture techniques)?		√	

5.	Will the study involve discussion of sensitive topics (e.g. sexual activity, drug use)?		√	
6.	Could the study induce psychological stress or anxiety or cause harm or negative consequences beyond the risks encountered in normal life?		√	
7.	Will blood or tissue samples be obtained from participants?		√	
8.	Are drugs, placebos or other substances (e.g. food substances, vitamins) to be administered to the study participants or will the study involve invasive, intrusive or potentially harmful procedures of any kind?		√	
9.	Is pain or more than mild discomfort likely to result from the study?		√	
10.	Will the study involve prolonged or repetitive testing?		√	
11.	Will it be necessary for participants to take part in the study without their knowledge and consent at the time (e.g. covert observation of people in non-public places)?		√	
12.	Will financial inducements (other than reasonable expenses and compensation for time) be offered to participants?		√	
13.	Is there any possible risk to the researcher (e.g. working alone with participants, interviewing in secluded or dangerous)?		√	

If you have ticked 'no' to all questions attach the completed and signed form to your project approval form and send to your programme/project co-ordinator for their records. Undergraduate and MA/MSc students should retain a copy of the form and submit it with their research report or dissertation (bound in at the end). MPhil/PhD students should submit a copy to the Faculty Research Degrees Sub-Committee with their application for registration (RD1) and forward a copy to their Faculty Ethics Committee. Members of staff should send a copy to their Faculty Ethics Committee.

If you have ticked 'yes' to ***any*** of the questions, please describe the ethical issues raised on a separate page. You will need to submit your plans for addressing the ethical issues raised by your proposal using the 'Application for Ethical Approval' form which should be submitted to the relevant Faculty Ethics Committee. This can be obtained from the University website (http://www.rdu.mmu.ac.uk/ethics/index.php).

If you answered yes to question 1, you may also have to submit an application to the appropriate external health authority ethics committee, and send a copy to the Faculty Ethics Committee for their records.

Please note that it is your responsibility to follow the University's Guidelines on Good Research Practice and any relevant academic or professional guidelines in the conduct of your study. **This includes providing appropriate information sheets and consent forms, and ensuring confidentiality in the storage and use of data.** Any significant change in the question, design or conduct over the course of the research should be notified to the relevant committee and may require a new application for ethics approval.

Approval for the above named proposal is granted

I confirm that there are no ethical issues requiring further consideration. *(Any subsequent changes to the nature of the project will require a review of the ethical consideration.)*

Signature of Supervisor: ________________________ Date: ____________

Approval for the above named proposal is not granted

I confirm that there are ethical issues requiring further consideration and will refer the project proposal to the appropriate Committee** Signature of Supervisor: _________________________ Date: ____________

** In accordance with Faculty/Department procedures

Appendix vi

November 2011

Manchester Metropolitan University

Research Study Information for Parents

This study aims to develop a British assessment of inferential comprehension for primary-school children.

I would like to tell you about some research that I am asking your child to take part in. This project has been approved by Manchester Metropolitan University's (MMU) Academic Ethics Committee.

What is the project about?

I am trying to find out about primary school children's understanding of complex spoken language. I (and some student- or graduate- Speech and Language Therapists) will be assessing as many 5-11 year-old children as possible from each year group in the Autumn, Spring and Summer terms, 2011 – 2012.

What will happen if my child takes part in the project?

I will see your child for one session of about 30 minutes at school and ask him/her a series of questions. I will write down your child's answers. I will audio- tape a small number of children so that my supervisors can check if my scoring system works. I will only do this if you have given specific consent for your child. When your child is finished, I will thank him/her and give a sticker.

The only information I need about your child is their school year and the month and year of their birth. If you agree to their participation, this information will be provided by school. Your child's full name will NOT be put on the assessment sheets. Any audio-recordings of your child will be kept in a locked secure filing cabinet and will be destroyed once the PhD has been completed.

What will happen if my child does not take part in the project?

There are no consequences for not allowing your child to participate in this study and you can withdraw your child at any time. However, I do hope you will agree to their inclusion.

Who is in charge of the project?

The study is being carried out by Anne Hewitt. I am a Senior Speech and Language Therapist (SLT) with Sefton PCT and have worked with children for 21 years. I am also a part-time senior lecturer at MMU. I have a current enhanced CRB check with Sefton PCT. I work in one of the units in the school every week so am already a familiar face to the children in the school. Any students or graduate SLTs who help out with the assessments will all have a current CRB check.

What happens with the work that my child does?

The results of the project will help me to develop an assessment of advanced understanding in primary-aged children.

What if I have questions about the project?

If you would like to ask any questions or get more information about the project, I would be happy to speak with you. Please contact me using the details at the bottom of this information sheet.

Thank you for your time.

Anne Hewitt, M.Ed., Senior Speech & Language Therapist,
Senior Lecturer in Speech Pathology & Therapy
Manchester Metropolitan University

Address:

e-mail:

Please fill in your child's name and return this slip to your child's teacher ONLY if you would NOT like your child to participate in the study.

Research study for the development of a British assessment of inferential comprehension for primary-school children.

I would **NOT** like____________________________(child's name) to participate in the research project

Class ______________

Signed ______________________(parent/guardian)   Date ______________

Please fill in your child's name and return this slip to your child's teacher if you give permission for me to audio-tape your child's responses.

Research study for the development of a British assessment of inferential comprehension for primary-school children.

I DO WANT my child __________________ (child's name) to be included in the research project and I give my permission for his/her responses to be audio-taped.

Class ___________

Signed ___________________________ (parent/guardian)   Date ___________

Appendix vii Emotions pictures

Appendix viii

Yes

Maybe

No

Appendix ix

HICIT Pilot Assessment
(Hewitt Inferential Comprehension and Idioms Test (A.H. 2010)
Pilot version (5-10 years)

Date(s) assessment trialled: ___________

Name of assessor __________________

Child's initials: ________ **Child's sex:** **M / F**

Age of child when assessed in years and months: ___________

Initial instructions to child: "You can ask me to repeat any questions as many times as you want. You probably won't know all the answers and it is fine to say 'Don't know' if you don't know an answer."

1) Simple Deductions **Score (/10)**

1. I am a fruit. I can be red or green. I am round. I am crunchy. I grow on a tree.
   What am I? (apple)
2. I am an animal. I am pink or brown. I live on a farm in a sty. People say I am greedy. I have a curly tail.
   What am I? (pig)
3. I am clothes. I am often made from wool. You put me on your hands. I keep your hands warm when it is cold.
   What am I? (gloves)
4. I am a musical instrument. You bang me with your hand or with sticks. I make a loud noise.
   What am I? (drum)
5. I am up in the sky. I am bright. I shine on you. I make you hot in the summer.
   What am I? (sun)

6. I am food. I have a hard shell. You sometimes need to crack my shell to get at me. There are lots of different kinds of me. Squirrels like to eat me.
   What am I? (nut)
7. I am made of metal and waterproof material. I have a handle. You put me up and down. I keep you dry in the rain.
   What am I? (umbrella)
8. I am an animal. You would see me in a zoo. I look like a horse. I have black and white stripes.
   What am I? (zebra)
9. I am a musical instrument. I am made of wood. I have strings. You play me with a bow. You hold me under your chin.
   What am I? (violin)
10. You find me inside and outside buildings. I have a frame. You can see through me. If you throw a hard ball at me I could smash.
    What am I? (window)

2) Why Question Inferences **Score (/20)**

a) Explaining obstacles to action and experience

1. Why can't you read in the dark?

2. Why can't you make a snowman in the summer time?

3. Why is it silly to go shopping without any money?

4. Why is it silly to hang washing out to dry when it is raining?

5. Why shouldn't you shout in a library?

b) Cause-effect/ Prevention

6. Why mustn't you play with matches?

7. Why should you wear an apron or old clothes when you are painting?

8. Why do people have umbrellas?

9. Why do you need a towel when you go swimming?

10. Why do we need pedestrian crossings on roads?

c) Justifying a decision

11. Why do people have jobs?

12. Why should you put the lid back on paint when you've used it?

13. Why would you decide you need to wear a coat to go outside?

14. Why shouldn't you agree to take a lift in a car from people you don't know?

15. Why do we wear shoes?

d) Construction of objects

16. Why are windows made out of glass not bricks?

17. Why are wellies made out of rubber or plastic not paper?

18. Why are mirrors made out of shiny surfaces not cardboard?

19. Why is a pan made out of metal not chocolate?

20. Why are teeth covered in enamel not cotton wool?

3) Making Predictions Inferences **Score (/8)**

1. What would happen to a flower if it didn't get any water?

2. What would happen to an egg if you dropped it on the floor?

3. What would happen if you had a glass of water full up to the top and then you added 6 ice cubes?

4. What would happen if you left a block of ice in the sun?

5. What would happen if you took away the bottom tin in a stack of tins at the supermarket?

6. What would happen if cars didn't stop at red traffic lights?

7. What would happen if you were very late for school every day?

8. If I showed you a picture in a book and then closed the book, would you still be able to see the picture?

<u>4. Formulating Solutions Inferences (What could you do?)</u> Score /21

1. What could you do when you are hungry?

2. What could you do when you are cold?

3. What could you do when you are dirty?

4. What could you do if you have lost something?

5. What could you do if you accidentally broke something that you had borrowed from a friend?

6. What could you do if you wanted to eat your soup but it was too hot?

7. What could you do if you forgot to take your PE kit to school on a day you had PE?

8. What could you do if you burnt a cake you had baked for your mum's birthday?

9. What could you do if you wanted to buy a toy but you didn't have enough money?

10. What could you do if you got locked inside the bathroom?

11. What could you do if you lost a letter the teacher had given you to take home to your parents?

12. What could you do if you fell over and hurt yourself?

13. What could you do if you saw a cat stuck in a tree?

14. What could you do if you locked yourself out from your house and nobody else was at home?

15. What could you do if you saw some children bullying another child at school?

16. What could you do if you dropped a glass of milk all over your kitchen floor?

17. What could you do if you found a purse that somebody had dropped?

18. How could you talk to somebody who lives hundreds of miles away from you?

19. How could you get to the top floor of a very tall block of flats?

20. How could you find out what your teacher looked like when s/he was a little girl/boy?

21. How do we keep milk fresh and cold?

<u>5. Explaining Inferences (How do you know?)</u> **Score (/22)**

1. How do you know that someone is tired?

2. How do you know that someone is too hot?

3. How do you know that someone is thirsty?

4. How do you know that someone is scared?

5. How do you know that someone is angry?

6. How do you know that someone has got toothache?

7. How do you know that someone is a policeman?

8. How do you know what's inside a box of something at the supermarket?

9. How do you know that someone doesn't like the dinner their mum gives them?

10. How do you know that someone thinks you are playing your music too loudly?

11. How do you know that someone has been eating biscuits in the kitchen?

12. How do you know that it is raining outside when someone comes in the front door?

13. How do you know that someone doesn't want to play with you?

14. How do you know that someone doesn't want you to hear what they are saying to someone else?

15. How do you know that someone thinks there is a bad smell in the room?

16. How do you know that your neighbours are moving house?

17. How do you know that a car has broken down?

18. How do you know that someone has put too much bubble bath in the bath?

19. How do you know that someone has walked through the house with muddy shoes?

20. How do you know that someone is having a birthday party?

21. How do you know that someone has got a hole in their shopping bag?

22. How do you know that a group of children want you to come and join in their game?

6) Making Inferences from short passages **Score (/16)**

1. Ellie paid the money to get in. She got changed and put her clothes in a locker. She put the key band on her wrist and took her towel with her.
What was Ellie doing? (going swimming)

2. Dan's dad took him to somewhere outside. He pushed Dan up and down on something and pushed him round and round on something else. Dan climbed up some steps and went down something.
Where were Dan and his dad? (at the park)

3. Nadia wanted to listen to her ipod but when she turned it on nothing happened.
Why?

4. Raj was playing on the beach. He trod on something and had to go to hospital?
What did Raj tread on?

5. Dad bought a chicken to cook for Sunday dinner but the family ended up eating just vegetables and gravy.
What happened?

6. The Polanski family went out for a walk in the countryside. They came back home very soon after they went out.
Why?

7. Zak and his friend Rashid went out for a bike ride. They came back later than they were supposed to and Zak was pushing his bike.
Why?

8. Alisha and her mum went to see a concert but had to come back home without seeing it.
What happened?

9. Katy went to play at her friend Rebecca’s house but she came home early.
Why?

10. James and his dad are sitting in the front row. The curtain goes up and people in costumes come on and talk loudly and sing.
Where are James and his dad?

11. Mum and dad have just eaten a nice meal in an expensive restaurant. They ask for the bill and then both look worried and embarrassed when it comes.
Why?

12. A group of friends went out to the cinema together. They had planned to get the bus home afterwards but ended up taking a taxi.
Why?

13. Mum and dad went to do some shopping but they came back to the house with nothing.
Why?

14. Jamal ordered a game from an internet company but it never arrived.
Why?

15. Jasmine is leaving her house. Her mum says ‘Don’t forget your packed lunch and your homework.’
Where is Jasmine going?

16. Dad hung the washing out on the line but when he went to bring it back in it wasn't there.
What had happened?

<u>7) Taking others' perspective inferences</u> **Score (/8)**

1. If Jenny was hungry would she finish her dinner?

2. If Tom was cold, would he take his jumper off?

3. If Billy hurt his friend would his friend be happy?

4. Would a fish be able to fly?

5. What would you do if your friend got stuck in some mud?

6. What would you do if your friend was crying?

7. Poppy was the only girl in her class who didn't get a valentine's day card. How did she feel?

8. Sam and Ruby had a date to go to the cinema but Ruby didn't turn up. How did Sam feel?

Instructions to the child: "For the next section, choose the feeling that you think fits best from: happy; sad; angry; or scared."

<u>8) Situation-based emotions inferences</u> **Score (/16)**

1. Rachel's pet cat died. How does she feel? (sad)
2. A boy took a toy off Jake. How does Jake feel? (angry)
3. A big fierce dog chased Raj down the road. How does he feel? (frightened)
4. Lucy's dad bought her some sweets. How does she feel? (happy)
5. Alice was invited to a friend's birthday party. How does she feel? (happy)

6. Jamelia painted a lovely picture but a girl came and scribbled on it on purpose. How does Jamelia feel? (angry)
7. Stefan broke his favourite toy. How does he feel? (sad)
8. Three-year old Tom has lost his mum in the shopping centre. How does he feel? (frightened)
9. A boy hits Gemma's sister at school. How does Gemma feel? (angry)
10. Jamie's dad takes him for tea at McDonalds but McDonalds is closed. How does Jamie feel? (sad)
11. Rahal's friend has brought him a present back from his holiday. How does Rahal feel? (happy)
12. Katy is alone in a dark, empty house. How does she feel? (frightened)
13. Laura got told off by the teacher but it was another child's fault. How does Laura feel? (angry)
14. Jake has won first prize in a competition. How does he feel? (happy)
15. Sophie is stuck in a tree, about to fall. How does she feel? (frightened)
16. Shazia's family holiday is cancelled at the last minute. How does Shazia feel? (sad)

Instructions to the child: "For the next section, choose the feeling that you think fits best from: happy or sad"

<u>9) Desire-based emotions inferences</u> **Score (/8)**

1. Sam wants to go on the train. Sam and his dad go on the train. How does Sam feel? (happy)
2. Sam wants to go on the train. Sam and his dad go in the car. How does Sam feel? (sad)
3. Lucy wants to go to the park. Lucy's sister takes her to the supermarket. How does Lucy feel? (sad)
4. Lucy wants to go to the park. Lucy's sister takes her to the park. How does Lucy feel? (happy)
5. Raj wants a car book. His mum buys him a train book. How does Raj feel? (sad)
6. Raj wants a car book. His mum buys him a car book. How does he feel? (happy)
7. Holly wants a banana in her lunch box. Holly's dad puts a banana in her lunch box. How does Holly feel? (happy)

8. Holly wants a banana in her lunch box. Holly's dad puts an apple in her lunch box. How does Holly feel? (sad)

Instructions to the child: "For the next section, choose the feeling that you think fits best from: happy or sad"

<u>10) Belief-based emotions inferences</u> **Score (/8)**

1) Matthew wants fish and chips for tea and he will be getting fish and chips.
Matthew thinks there is fish and chips for tea.
1a) How does he feel? (happy).
1b) His mum gives him fish and chips for tea. How does he feel? (happy)

2) Raj wants a toy plane for his birthday but he is getting a teddy.
Raj thinks he is getting a toy plane for his birthday.
2a) How does he feel? (happy).
2b) He gets teddy for his birthday. How does he feel? (sad)

3) Jane wants to go to the beach but she is going to the library.
Jane thinks her dad is taking her to the library.
3a) How does she feel? (sad).
3b) Her dad takes her to the library. How does she feel? (sad)

4) Samia wants to go clothes shopping with her mum and she will be going clothes shopping.
Samia thinks she is going to help her mum clean the house.
4a) How does she feel? (sad).
4b) Mum takes her clothes shopping. How does she feel? (happy)

<u>11. Mental state verb inferences</u> **Score (/14)**

I am going to tell you about 14 different people and a door. If you think the person has definitely locked the door say 'yes'. If you think they might or might not have locked the door say 'maybe'. If you think the person has definitely NOT locked the door say 'no'.

1. Karen knows that the door is locked. Is the door locked? (yes)
2. Jack thinks that the door is locked. Is the door locked? (maybe)

3. Raj manages to lock the door. Is the door locked? (yes)
4. Shazia forgets to lock the door. Is the door locked? (no)
5. James wants to lock the door. Is the door locked (maybe)
6. Jasmine realises that she locked the door. Is the door locked (yes)
7. Zak is sorry that he locked the door. Is the door locked? (yes)
8. Alisha promises to lock the door. Is the door locked? (maybe)
9. Dan remembers to lock the door. Is the door locked? (yes)
10. Katy agrees to lock the door. Is the door locked? (maybe)
11. Jamal learns that the door is locked. Is the door locked? (yes)
12. Ellie believes she has locked the door. Is the door locked? (maybe)
13. Billy is happy that he locked the door. Is the door locked? (yes)
14. Sue is careful to lock the door. Is the door locked? (yes)

12) Strange Stories **Score (/16)**

NB. If the child answers 'yes' to the 'Is it true?' question probe 'why?

1. Pretend

John and Rashid are playing spacemen. John picks up a bowl and puts it on his head. He says: 'This bowl is a space helmet.'

1a) Is it true what John says? (No)
1b) Why does he say it? (Pretending)

2. Lying

Julie hates going to the doctor. She has very bad stomach pains and is holding her stomach. Her mum asks "Have you got tummy ache?" Julie says 'No'.
2a) Is it true what Julie says? (No)
2b) Why does she say it? (lying so she doesn't have to go to the doctor's)

3. Deception

Jane's sister Holly always tells lies to Jane and Jane knows she does. Holly has hidden Jane's hairdryer under her bed. Jane asks: "Have you hidden my hairdryer under your bed or in your drawer?" Holly says "It's in my drawer".

3a) Is it true what Holly says? (No)

3b) Why (because it is under her bed; because she always tells lies to Jane)

3c) Where will Jane look for her hairdryer? (under the bed)

If the child gives the wrong answer to the above say: ‘I think she will look under the bed.’ Then ask ‘Why?’
3d) Why? (Because she knows Holly tells lies so will do the opposite of what she says).

<u>4. Misunderstanding</u>
Jack hurt another child on purpose at school, but no teachers saw him do it. At the end of the day he started to walk home. A teacher ran after him shouting ‘Stop Jack!’ Jack turned round and said ‘I’m really sorry Miss’. The teacher only wanted to give him his lunch box that he had forgotten.

4a) Why did Jack say ‘I’m really sorry Miss’ (because he thought the teacher was going to tell him off for hurting the other child).

4b) Did the teacher want to tell Jack off? (No)

4c) How did the teacher feel when Jack said ‘I’m really sorry Miss’ (surprised, puzzled, confused)

If the child gives the wrong answer to the above say: ‘I think she was surprised.’ Then ask ‘Why?’
4d) Why? (because she was just giving him his lunch box).

<u>5.) Contrary Emotions</u>
James’ group of friends is going swimming on Saturday morning. James really wants to go with them but he knows that Sam, a big boy who bullies him at school goes swimming in the same pool on Saturday mornings. When his friends ask him if he wants to go swimming with them he says ‘No.’

5a) Is it true what James says? (No) Can also accept Yes for this one.

5b) Why does he say he doesn't want to go swimming? (to avoid the bully)

<u>6. White lie</u>
Tom's friend Samia has just had her hair cut. Tom doesn't like it. He doesn't think it suits her. He liked her old style better. When Samia asks Tom 'Do you like my new hair style? Tom says 'Yes, it's very nice.'

6a) Is it true what Tom said? (No)
6b) Why did he say it? (lied so as not to hurt Samia's feelings)

<u>13) Faux pas</u> **Score (/18)**

1. Anne introduces Bob to her French cousin Marie. Kevin doesn't hear this introduction. He comes over and the conversation turns to different nationalities. Kevin says 'I hate the French.'

1a) Did Kevin know that Marie was French? (No)
1b) How does Marie feel (upset/ angry)
1c) How does Kevin feel when he finds out that Marie is French? (embarrassed, guilty)

2. Mrs Patel has Mrs Brown, an old friend, over for tea. She hasn't seen Mrs Brown for years. Mrs Brown accidentally knocks over a vase and breaks it. Mrs Patel says 'It doesn't matter. It was a wedding present that I never liked anyway. I can't remember who bought it for us." Mrs Brown says 'I bought it for you.'

2a) How does Mrs Brown feel? (upset/angry)
2b) Did Mrs Patel remember that Mrs Brown had bought her the vase? No
2c) How does Mrs Patel feel (embarrassed, guilty)

3. It was the school disco. Jill was in one of the toilets. Karen and Samia came in to the toilets afterwards. Karen said: 'Did you see how Jill was dancing. Wasn't she terrible!' Then Jill came out of the toilet.

3a) How does Karen feel? (embarrassed, guilty)

3b) How does Jill feel? (upset/angry)
3c) Did Karen know that Jill was in one of the toilets? (No)

4. Mike and Rashid are talking about different jobs. Rashid says ‘I hate taxi drivers. They’re all so common and rude.’ Mike says ‘My dad’s a taxi driver.’

4a) How does Rashid feel? (embarrassed, guilty)
4b) Did Rashid know that Mike’s dad was a taxi driver? (No)
4c) How does Mike feel? (upset/ angry)

5. Jake and Jenny are brother and sister. Jake has got very long hair. A friend of his mum’s comes to visit. She sees Jake and says ‘You must be Jenny, nice to meet you’. He says ‘I’m Jake, not Jenny.’

5a) How does Jake feel? (upset/ angry)
5b) How does the mum’s friend feel? (embarrassed/ guilty)
5c) Did the mum’s friend know that Jake was a boy when she first saw him? (No)

6. Laura worked as a clothes shop assistant. A lady with a fat tummy came in and tried on a dress. Laura said ‘That looks lovely and you will still be able to wear it when you have had the baby.’ The lady said ‘I’m not pregnant.’

6a) Did Laura know that the lady was not pregnant? (No)
6b) How does Laura feel? (embarrassed/ guilty)
6c) How does the lady feel? (upset/ angry).

Instructions to the tester: Say all of these idioms with a fairly neutral tone of voice and facial expression so you do not give away the meaning.

<u>14) Idioms</u> /25
What do the following sayings mean?

1. He’s full of beans today.

2. He’s driving me round the bend.

3. She got out of the wrong side of bed today.

4. Can you give me a hand?

5. Don’t let the cat out of the bag.

6. Keep your eyes peeled.

7. He’s having a whale of a time.

8. This is a piece of cake.

9. You’ve hit the nail on the head.

10. It goes in one ear and out the other.

11. It’s not my cup of tea.

12. I’ve got butterflies in my tummy.

13. Put a sock in it.

14. She’ll hit the roof.

15. He’s going to turn over a new leaf.

16. You’re pulling my leg.

17. I need to sleep on it.

18. I’m all ears.

19. He has a heart of gold.

20. Let’s hit the road.

21. They get on like a house on fire.

22. I’m tied up at the moment.

23. My lips are sealed.

24. Get your skates on.

25. She’s over the moon.

Anne Hewitt, Dec 2009:

Please return any completed assessment forms and comments to:

Appendix x

<u>Hewitt Inferential Comprehension and Idioms Test, 5-10 years</u>
<u>Final version June 2012</u>

Date assessment trialled: ____________
Name of assessor: __________________
Child's initials: ________ Child's sex: M / F
Child's Month and Year of Birth: __________ Year Group: ______
Age of child when assessed in years and months: ________________
Region of Great Britain the school is in: ______________________
Percentage of FREE SCHOOL MEALS in the school: ____________

<u>Instructions to testers</u>

Start by telling the child: 'I have got lots of questions to ask you. You may not know all the answers and it is fine to say 'I don't know' if you do not know the answer to any question'.

Say that some questions have more that one right answer.

Tell the older children (year 3 onwards) that a lot of the questions might be very easy for them.

Tell the child you can repeat each question as many times as he/she needs you to.

If the child gives an idiosyncratic response (e.g. Section 2, Q6 'Why can't you read in the dark?': 'Because I didn't eat my carrots', Q7 'Why are windows made out of glass and not out of bricks?': 'The Georgians had bricks in their windows to avoid paying window tax'; Section 4, Q11 'How could you find out what your teacher looked like when she/he was a little girl?: 'Go back in time'; Section 5, Q3 'How do you know that someone is angry?': 'In a cartoon they would have steam coming out of their ears') prompt them with 'Yes, but what about in real life/ nowadays?'

If you think **the child has partially answered a question**, prompt with ‘Anything else?’ NOT with ‘Why?
For example: Section 2, Q4 Tester ‘Why mustn’t you play with matches?’: Child ‘Because they’re dangerous’; Tester ‘Anything else?’

Section 5 ‘How do you know that....?’ If the child answers ‘Because they tell you’ for any of these questions prompt with ‘Yes, **and how else** do you know?’ Question 8. ‘How do you know that someone has got toothache?’ If the child replies ‘they go to the dentist’ repeat the question stressing the **you** (Yes, but how do YOU know they have got toothache?)

This is a **verbal test** so keep the test form out of sight as much as possible so the child does not try to read the questions.

Write down the child’s exact verbal response to each of the following questions. If he/she uses gesture to answer some questions describe the gestures used.

<u>1) Simple Deductions</u> **Score (/8)**

1. I am a musical instrument. You bang me with your hand or with sticks. I make a loud noise.
   What am I? (drum/ tambourine/ cymbal/ gong)
2. I am an animal. You would see me in a zoo. I look like a horse. I have black and white stripes.
   What am I? (zebra)
3. I am an animal. I am pink or brown. I live on a farm in a sty. People say I am greedy. I have a curly tail.
   What am I? (pig)
4. I am clothes. I am often made from wool. You put me on your hands. I keep your hands warm when it is cold.
   What am I? (gloves/ mittens/ muff)
5. I am food. I have a hard shell. You sometimes need to crack my shell to get at me. There are lots of different kinds of me. Squirrels like to eat me.
   What am I? (nut, any kind)

6. You find me inside and outside buildings. I have a frame. You can see through me. If you throw a hard ball at me I could smash.
What am I? (window)
7. I am a musical instrument. I am made of wood. I have strings. You play me with a bow. You hold me under your chin.
What am I? (violin/ viola/ fiddle)
8. I am made of metal and waterproof material. I have a handle. You put me up and down. I keep you dry in the rain.
What am I? (umbrella/ brolly)

2) Why Question Inferences **Score (/12)**

1. Why is it silly to hang washing out to dry when it is raining?

2. Why do you need a towel when you go swimming?

3. Why do people have jobs?

4. Why mustn't you play with matches?

5. Why are mirrors made out of shiny surfaces not cardboard?

6. Why can't you read in the dark?

7. Why are windows made out of glass not bricks?

8. Why is a pan made out of metal not chocolate?

9. Why are wellies made out of rubber or plastic not paper?

10. Why shouldn't you agree to take a lift in a car from people you don't know?

11. Why are teeth covered in enamel not cotton wool?

(NB If the child asks what enamel is you can say: 'The stuff on the outside of teeth')

12. Why do we need pedestrian crossings on roads?
(NB If the child asks what a pedestrian crossing is you can say: 'Like a zebra crossing')

<u>3) Making Predictions Inferences</u> **Score (/6)**

1. What would happen if you left a block of ice in the sun?

2. What would happen if you were very late for school every day?

3. What would you do if your friend got stuck in some mud?

4. Sam and Ruby had a date to go to the cinema but Ruby didn't turn up. How did Sam feel?

5. What would happen if you took away the bottom tin in a stack of tins at the supermarket?

6. What would happen if you had a glass of water full up to the top and then you added 6 ice cubes?

<u>4. Formulating Solutions Inferences (What could you do?)</u> **Score /12**

1. What could you do when you are cold?

2. What could you do if you wanted to buy a toy but you didn't have enough money?

3. What could you do when you are hungry?

4. What could you do if you wanted to eat your soup but it was too hot?

5. How could you get to the top floor of a very tall block of flats?

6. What could you do if you dropped a glass of milk all over your kitchen floor?

7. What could you do if you burnt a cake you had baked for your mum's birthday?

8. How do we keep milk fresh and cold?

9. What could you do if you accidentally broke something that you had borrowed from a friend?

10. How could you talk to somebody who lives hundreds of miles away from you?

11. How could you find out what your teacher looked like when s/he was a little girl/boy?

12. What could you do if you locked yourself out from your house and nobody else was at home?

<u>5. Explaining Inferences (How do you know?)</u> Score (/12)

N.B Prompt all 'Because they tell you' answers in this section.

1. How do you know that someone is tired?

2. How do you know that a group of children want you to come and join in their game?

3. How do you know that someone is angry?

4. How do you know that someone has been eating biscuits in the kitchen?

5. How do you know that someone is too hot?

6. How do you know that someone has put too much bubble bath in the bath?

7. How do you know that your neighbours are moving house?

8. How do you know that someone has got toothache?

9. How do you know what's inside a box of something at the supermarket?

10. How do you know that someone doesn't want you to hear what they are saying to someone else?

11. How do you know that someone thinks there is a bad smell in the room?

12. How do you know that someone thinks you are playing your music too loudly?

<u>6) Making Inferences from short passages</u> **Score (/10)**

1. Dad hung the washing out on the line but when he went to bring it back in it wasn't there.
What had happened?

2. Raj was playing on the beach. He trod on something and had to go to hospital?
What did Raj tread on?

3. Ellie paid the money to get in. She got changed and put her clothes in a locker. She put the key band on her wrist and took her towel with her.
What was Ellie doing?

4. Alisha and her mum went to see a concert but had to come back home without seeing it.
What happened?

5. Dan’s dad took him to somewhere outside. He pushed Dan up and down on something and pushed him round and round on something else. Dan climbed up some steps and went down something.
Where were Dan and his dad?

6. A group of friends went out to a party together. They had planned to get the bus home afterwards but ended up taking a taxi.
Why?

7. Jane and her friend Sasha went out for a bike ride. They came back later than they were supposed to and Jane was pushing her bike.
Why?

8. James and his dad are sitting in the front row. The curtain goes up and people in costumes come on and talk loudly and sing.
Where are James and his dad?

9. Mum and dad have just eaten a nice meal in an expensive restaurant. They ask for the bill and then both look worried and embarrassed when it comes.
Why?

10. Dad bought a chicken to cook for Sunday dinner but the family ended up eating just vegetables and gravy.
What happened?

7) Situation-based emotions inferences **Score (/16)**

Use the 4 emotions symbols for this section. Before you start check that the child knows what picture depicts what emotion by asking him/her to point to each one as you say them.
Instructions to the child: “For the next section, choose the feeling that you think fits best from: happy; sad; angry; or scared. There might be two

possible answers for some of these but choose the one feeling that fits the best."

1. Rachel's pet cat died. How does she feel? (sad)
2. A boy took a toy off Jake. How does Jake feel? (angry)
3. A big fierce dog chased Raj down the road. How does he feel? (scared)
4. Lucy's dad bought her some sweets. How does she feel? (happy)
5. Alice was invited to a friend's birthday party. How does she feel? (happy)
6. Jamelia painted a lovely picture but a girl came and scribbled on it on purpose. How does Jamelia feel? (angry)
7. Stefan broke his favourite toy. How does he feel? (sad)
8. Three year-old Tom has lost his mum in a busy shopping centre. How does he feel? (scared)
9. A boy hits Gemma's sister at school. How does Gemma feel? (angry)
10. Jamie's dad takes him for tea at McDonalds but McDonalds is closed. How does Jamie feel? (sad)
11. Rahal's friend has brought him a present back from his holiday. How does Rahal feel? (happy)
12. Katy is alone in a dark, empty house. How does she feel? (scared)
13. Laura got told off by the teacher but it was another child's fault. How does Laura feel? (angry)
14. Jake has won first prize in a competition. How does he feel? (happy)
15. Sophie is stuck in a tree, about to fall. How does she feel? (scared)
16. Shazia's family holiday is cancelled at the last minute. How does she feel? (sad)

8) Belief-based emotions inferences **Score (/8)**

Just have the 2 symbols for happy and sad visible for this section (i.e fold the emotions sheet in half).
Instructions to the child: "For the next section, choose the feeling that you think fits best from: happy or sad"

Matthew wants fish and chips for tea and he will be getting fish and chips.

Matthew thinks there are fish and chips for tea.
1) How does he feel? (happy).
2) His mum gives him fish and chips for tea. How does he feel? (happy)

Raj wants a toy plane for his birthday but he is getting a teddy.
Raj thinks he is getting a toy plane for his birthday.
3) How does he feel? (happy).
4) He gets a teddy for his birthday. How does he feel? (sad)

Jane wants to go to the beach but she is going to the library.
Jane thinks her dad is taking her to the library.
5) How does she feel? (sad)
6) Her dad takes her to the library. How does she feel? (sad)

Samia wants to go clothes shopping with her mum and she will be going clothes shopping.
Samia thinks she is going to help her mum clean the house.
7) How does she feel? (sad)
8) Mum takes her clothes shopping. How does she feel? (happy)

<u>9. Mental state verb inferences</u> **Score (/12)**

Use the page with the words Yes, Maybe and No on it. Point to the appropriate word as you give the instructions below.

'I am going to tell you about 12 different people and a door. If you think the door is definitely locked say 'yes'. If you think the door might or might not be locked say 'maybe'. If you think the door is definitely NOT locked say 'no'.

1. Karen knows that the door is locked. Is the door locked? (yes)
2. Jack thinks that the door is locked. Is the door locked? (maybe)
3. Raj manages to lock the door. Is the door locked? (yes)
4. Shazia forgets to lock the door. Is the door locked? (no)
5. James wants to lock the door. Is the door locked (maybe)
6. Jasmine realises that she locked the door. Is the door locked (yes)

7. Zak is sorry that he locked the door. Is the door locked? (yes)
8. Alisha promises to lock the door. Is the door locked? (maybe)
9. Dan remembers to lock the door. Is the door locked? (yes)
10. Jamal learns that the door is locked. Is the door locked? (yes)
11. Ellie believes she has locked the door. Is the door locked? (maybe)
12. Billy is happy that he locked the door. Is the door locked? (yes)

<u>10) Strange Stories</u> **Score (/12)**

NB. If the child answers 'yes' to the 'Is it true?' question probe 'why?

<u>Pretend</u>

John and Rashid are playing spacemen. John picks up a bowl and puts it on his head. He says: 'This bowl is a space helmet.'

1) Is it true what John says?
2) Why does he say it?

<u>Lying</u>

Julie hates going to the doctor. She has very bad stomach pains and is holding her stomach. Her mum asks "Have you got tummy ache?" Julie says 'No'.

3) Is it true what Julie says?
4) Why does she say it?

<u>Misunderstanding</u>

Jack hurt another child on purpose at school, but no teachers saw him do it. At the end of the day he started to walk home. A teacher ran after him shouting 'Stop Jack!' Jack turned round and said 'I'm really sorry for doing it Miss'. The teacher only wanted to give him his lunch box that he had forgotten.

5) Why did Jack say 'I'm really sorry for doing it Miss'

6) Did the teacher want to tell Jack off?

7) How did the teacher feel when Jack said ‘I’m really sorry for doing it Miss’

(If the child gives the wrong answer to the above say: ‘I think she was surprised.’ Then ask ‘Why?’)
8) Why?

Contrary Emotions
James’ group of friends is going swimming on Saturday morning. James really wants to go with them but he knows that Sam, a big boy who bullies him at school goes swimming in the same pool on Saturday mornings. When his friends ask him if he wants to go swimming with them he says ‘No.’

9) Is it true what James says?
10) Why does he say he doesn’t want to go swimming?

White lie
Tom’s friend Samia has just had her hair cut. Tom doesn’t like it. He doesn’t think it suits her. He liked her old style better. When Samia asks Tom ‘Do you like my new hair style? Tom says ‘Yes, it’s very nice.’

11) Is it true what Tom said?
12) Why did he say it?

11) Faux pas **Score (/12)**

N.B It is OK to remind a child who specific people are if they ask, e.g: Q1- If the child asks ‘Who was Karen again?’ Tester: She’s the one who said: ‘Did you see how Jill was dancing?’

Disco
It was the school disco. Jill was in one of the toilets. Karen and Sara came in to the toilets afterwards. Karen said: ‘Did you see how Jill was dancing. Wasn’t she terrible!’ Then Jill came out of the toilet.

1) How does Karen feel?
2) How does Jill feel?
3) Did Karen know that Jill was in one of the toilets?

Taxi driver

Mike and Rashid are talking about different jobs. Rashid says ‘I hate taxi drivers. They’re all so common and rude.’ Mike says ‘My dad’s a taxi driver.’

4) How does Rashid feel?

5) Did Rashid know that Mike’s dad was a taxi driver?

6) How does Mike feel?

Clothes shop

Laura worked as a clothes shop assistant. A lady with a fat tummy came in and tried on a dress. Laura said ‘That looks lovely and you will still be able to wear it when you have had the baby.’ The lady said ‘I’m not pregnant.’

7) Did Laura know that the lady was not pregnant?

8) How does Laura feel?

9) How does the lady feel?

Vase

Mrs Patel has Mrs Brown, an old friend, over for tea. She hasn’t seen Mrs Brown for years. Mrs Brown accidentally knocks over a vase and breaks it. Mrs Patel says ‘It doesn’t matter. It was a wedding present that I never liked anyway. I can’t remember who bought it for us.” Mrs Brown says ‘I bought it for you.’

10) How does Mrs Brown feel?

11) Did Mrs Patel remember that Mrs Brown had bought her the vase?

12) How does Mrs Patel feel?

12) Idioms **Score (/20)**

Instructions to the tester: Say all of these idioms with a fairly neutral tone of voice and facial expression and do not gesture, so that you do not give away the meaning.

Introduce the section by giving the following example:
“Some sayings mean something different from their words. For example ‘give me a hand’ doesn’t mean ‘chop a hand off and give it to me’. It means ‘help me please.’ What do the following sayings mean? It is okay to say ‘don’t know’ if you do not know what it means.”

1. Keep your eyes peeled
2. I’ve got butterflies in my tummy
3. My lips are sealed
4. Let’s hit the road
5. Put a sock in it
6. This is a piece of cake
7. She’s over the moon
8. He’s having a whale of a time
9. She got out of the wrong side of bed today
10. He’s driving me round the bend
11. Get your skates on
12. It’s not my cup of tea
13. They get on like a house on fire
14. He’s going to turn over a new leaf
15. You’re pulling my leg
16. I’m tied up at the moment

17. She’ll hit the roof

18. I need to sleep on it

19. It goes in one ear and out the other

20. You’ve hit the nail on the head

Anne Hewitt, June 2012:
Please return any completed assessment forms and comments to:

Appendix xi

Final Marking Criteria for the HICIT

Please use your clinical discretion to mark the few answers that may not be covered by the marking criteria below. Mark any justified responses as correct.
For example:
Section 4, question 7: 'What could you do if you burnt a cake you had baked for your mum's birthday?'
Child: 'Give it to her because my mum likes burnt cake'

Also – accept local dialectal vocabulary. For example, 'narky' for 'moody'; 'get a wiggle on' for 'get your skates on.'

Acceptable answers

1) Simple Deductions

1. Drum/ tambourine/ cymbal/ gong
2. Zebra
3. Pig
4. Gloves/ mittens/ muff
5. Nut (any kind)
6. Window
7. Violin/ viola/ fiddle
8. Umbrella/ brolly

2) Why Questions Inferences

1. Because it'll get wet (again); it won't dry; it'll get (even) wetter.

2. To dry yourself (off); so you can get dry; so your clothes don't stick to you afterwards; because you don't want to go home in wet clothes.

3. To get/earn money; to pay the bills (including rent); to buy things for the family; so they can (afford to) buy things; so they can go on holiday; so they can have a better life.

4. Because you can burn yourself/others; It'll set fire/go on fire; You could burn the house down.

5. So you can see yourself/your reflection; cardboard doesn't reflect; so you can see what you look like.

6. Because you can't see the letters/words/writing/anything; It's too dark to see the words; You need a light so you can see; Because your eyes don't work in the dark; because words don't glow in the dark.

7. So we can see through them; To let light/sun in; Because you can't see out of/through bricks; Glass is see through/transparent (and bricks aren't).

8. Because it/the chocolate would melt; because metal doesn't melt.

9. Because rubber/plastic is waterproof/doesn't absorb water/stays dry; so water doesn't come through/into them; because paper goes soggy; because paper soaks up/absorbs water/gets wet/ Because water dissolves paper /Paper breaks in water/ water ruins paper/ your feet would get wet.

10. You shouldn't go with strangers/people you don't know; They might hurt/abduct/kidnap you; They might take you away/steal you/take you hostage; They might do bad things to you/ They might be nasty people/ They might trick you.

11. Because enamel is hard and cotton wool is too soft; enamel protects the tooth; cotton wool would go soft/dissolve/ so the teeth don't break.

12. So you can cross the road safely; so the cars know when to stop (to let the people cross); so people don't get hit by cars/run over.

3. Making Predictions Inferences

1. It would melt.

2. You would be told off/shouted at/get in trouble/be expelled; You would get a late mark/detention; You would miss a lot of school/ lessons/learning/everything; You wouldn't know anything; You'd have to sign in at the office; your parents would be informed/contacted/ telephoned.

3. Try to help him/pull him out; throw him a rope; get a stick and pull him out; go and get/phone for help; call the fire brigade/police.

4. Sad/upset/miserable/disappointed/hurt/angry/furious/unloved/lonely/ worried/ anxious/concerned/ annoyed/ let down.

5. They would all fall down/over/tumble down; the stack would collapse.

6. The water would overflow/spill over/go over the top/splash out/ go everywhere/all over.

4. Formulating Solutions Inferences (What could you do?)

1. Put (any variation of) warm clothes/extra layers on/ wrap up warm; Put the heating/fire on; get a blanket/cover; go inside; get in bed; Drink something hot.

2. Save up; buy a cheaper/different one; borrow/earn some money; ask for more pocket money; go home and get some more money; go to the bank/ get some out of the cash point; wait for Christmas/birthday; leave the shop; you couldn't buy it; borrow someone else's.

3. Eat something; ask for food/a snack; say 'I want food'; buy some food.

4. Blow on it; wait/ leave it for a while (for it to cool); put milk/cream/cold water in it; dip some bread into it; fan it (with your hand)/ put it by the (open) window; put it in the fridge/freezer.

5. Go up/climb the stairs/steps; go in the lift/elevator.

6. Clean/mop/wipe it up/tidy it up/sweep it up; tell/go and get someone.

7. Make/bake another one/ something else instead; start/do it again; buy one; fix/repair/salvage it; cut the burnt bits off/cover the burnt bits in icing; say sorry/apologise; you couldn't eat it; get someone else to make one for you/help you make one.

8. (Put it) in the fridge/refrigerator/freezer/pantry.

9. Say sorry/apologise; tell the truth/ be honest/ confess; buy them a new one; give them one (the same)/replace it; pay them/give them money; try to fix/glue it/put it back together; be honest; say/explain what happened/what you did.

10. On the phone/telephone; internet; webcam/ video chat; skype them; face time/ letter/ be a pen pal/ facebook/ e-mail/ twitter.

11. Ask her/him to show you/bring in a picture/photo/video/DVD; Ask her/him to tell you/describe (what they looked like); ask her/his parents.

12. Go to another family member's/your neighbour's/a friend's house; find the spare key; ring/call/text someone; wait outside until someone comes home; try the back door; climb in a window/ try and break in/ break a window and climb in; go to your parents' work place.

5. Explaining Inferences (How do you know?)

1. They yawn (and stretch); they close/shut their eyes; they look sleepy; put their head down on the table; they have bags under their eyes; they go to bed; they lie down.

2. They ask you (if you want) to come and play/join in; they wave/beckon you over (to their group); they come/run over to you.

3. They frown/pull a cross face/ have lines on their forehead; their eye brows go together; their face goes red; they clench their fists; they grit/gnash their teeth; they shout/scream; they stamp their feet; their voice gets louder; they take it out on someone near them.

4. There are fewer biscuits than before; the biscuit packet/tin/jar is (nearly) empty/has been left open; the biscuits are all gone; there are crumbs (on the plate/table/their faces); the biscuit tin has been moved; they look guilty.

5. They pant; they are red faced; they sweat; they fan themselves/ put a fan on; they take some clothes off; they open the window/ they're shiny on their head.

6. The bubbles are everywhere/overflowing/all over (the bathroom); there are too many/loads of bubbles; lots of bubbles are floating round the bathroom; it's too bubbly.

7. They have a 'for sale'/moving house/ sold sign up; there is a removal van outside the house; they are putting all the stuff from their house into a van/lorry; there are lots of boxes; removal men are packing up their stuff.

8. They are holding/clutching their mouth/jaw/cheek; they keep their hand on their mouth/tooth; they are saying 'ow'/complaining and holding their mouth; they act in pain/ screaming with pain.

9. There will be a label/sign/picture/writing; it will say (on the box/ cover); it has it on the box; you could ask someone who works there.

10. They whisper; cover their mouth with their hand; put their hand over their mouth; they say 'Hold on a minute'; they say 'I just need to talk to them for a sec'; they say 'We want to talk in private/privately; they go in a huddle; they move/run away from you; they go off somewhere/ hide from you; they tell you it's a secret; they tell you to go away.

11. They hold their nose (with their fingers); they cover their nose up; they say 'phwoar'/ 'yuk'; they waft their hand; they put a peg on their nose; they say 'What's that (horrible/nasty) smell?/ They pull a face./ They spray/plug in an air freshener.

12. They cover their ears up/put their fingers in their ears; they come and tell you/shout at you to turn it down/off; they knock at your door/bang on your wall (and tell you to turn it down); they tell you; they pull the plug out.

6. Making inferences from short passages

1. It had blown away; someone else had brought it in; somebody stole/pinched/nicked it; it fell on the ground; an animal stole it/ bird took it for its nest.

2. Anything sharp or poisonous e.g.: A crab; (hard) shell; piece of glass; a spike/ thorn; a piece of (sharp) metal; a (sharp) stone; a jellyfish; a broken bottle; a (sharp) stick; a nail/screw/pin; a bee/wasp; a knife; a shark's tooth; a sea urchin.

3. Going swimming/ in the pool; going to the gym

4. They couldn't get a ticket; they lost their tickets; the concert was cancelled/called off; it was sold out; the concert had already finished/they missed it; there was a fire at the concert; they didn't take enough/any money for the tickets; they were too late to get in; they went on the wrong day; 1 of them (or the band members) were sick/ hurt/injured; they got stuck in traffic/delayed/ their car broke down; Alisha was too young to get in; there were too many people there; bad weather prevented it (e.g. a storm).

5. At the (swing) park/ playground; at a water park/ fun fair

6. They missed/were too late for the (last) bus/the buses didn't run that late; the bus didn't come/turn up/ stop; the bus had broken down/run out of petrol; the bus was full; they couldn't find the bus stop/station; the bus stop was too far away from where they were; they needed to get home quickly; a taxi was cheaper (for that many people); because they were too drunk; the bus driver didn't turn up for work/was ill.

7. Because her bike had a puncture/flat tyre/burst tyre/the tyre went down; her bike had broken; the chain came off/ broke; she had fallen off her bike; she had crashed; 1 of them had had an accident; she was too tired; her legs were tired; her feet/knees were sore.

8. At the theatre/ a show/ pantomime/ west end/ a play/ the opera house/ the circus/ a festival.

9. Because it was too expensive/cost too much/ they couldn't afford it/ they've overspent; they didn't have enough money to pay for it; they didn't realise how much it was going to cost; they couldn't pay; they forgot their money/wallet/purse.

10. The chicken was burnt/ overcooked/ set on fire; the chicken was off/stale/ mouldy/rotten/ out of date; Dad dropped the chicken; somebody/an animal stole/ate the chicken; he forgot to put the oven on; the chicken was raw; they decided to become vegetarians; Dad didn't know how to cook the chicken.

7) Situation-based emotions inferences

1. Sad

2. Angry
3. Scared
4. Happy
5. Happy
6. Angry
7. Sad
8. Scared
9. Angry
10. Sad
11. Happy
12. Scared
13. Angry
14. Happy
15. Scared
16. Sad

8) Belief-based emotions inferences

1. Happy
2. Happy
3. Happy
4. Sad
5. Sad
6. Sad
7. Sad
8. Happy

9. Mental state verb inferences

1. Yes
2. Maybe
3. Yes
4. No
5. Maybe
6. Yes
7. Yes
8. Maybe
9. Yes
10. Yes
11. Maybe
12. Yes

10. Strange Stories

1. No

2. Pretending/ playing at being spacemen/ using his imagination/ it could be a space hat in the game.
3. No
4. Lying so she doesn't have to go to the doctor's; she doesn't want her mum to know she is in pain; because she hates going to the doctor's.
5. Because he thought the teacher was going to tell him off for hurting the other child; He thinks the teacher's found out/knows what happened; he thinks the child he hurt has told the teacher.
6. No
7. Surprised/ puzzled/ confused.
8. Because she was just giving him his lunch box; She didn't know / she wondered what had happened/that he had hurt another child/ what he was talking about/ why he was saying sorry.
9. No. You can accept yes for this one too.
10. To avoid the bully.
11. No
12. He lied so as not to hurt her feelings/offend/disappoint her; to make her feel happy; told a white lie.

11. Faux Pas

1. embarrassed/ guilty/ ashamed/ full of shame/ awful/ terrible/ regretful/ wanting the ground to swallow them up/ bad/ mean/ rude/ sorry /wanting to take back what they said/ worried that Jill might have heard.
2. upset/ angry/ mad / annoyed /sad/ disappointed/offended/ shocked/ grumpy.
3. No
4. embarrassed/ guilty/ ashamed/ full of shame/ awful/ terrible/ regretful/ wanting the ground to swallow them up/ bad/ mean/ rude/ sorry/ wanting to take back what they said.
5. No
6. upset/ angry/ mad / annoyed/ sad/ disappointed/offended/ shocked/ grumpy.
7. No
8. embarrassed/ guilty/ ashamed/ full of shame/ awful/ terrible/ regretful/ wanting the ground to swallow them up/ bad/ mean/ rude/ sorry/ wanting to take back what they said.
9. upset/ angry/ mad / annoyed/ sad/ disappointed/offended/ shocked/grumpy/ embarrassed/ depressed
10. upset/ angry/ mad / annoyed/ sad/ disappointed/offended/ shocked/ grumpy
11. No
12. embarrassed/ guilty/ ashamed/ full of shame/ awful/ terrible/ regretful/ wanting the ground to swallow them up/ bad/ mean/ rude/ sorry wanting to take back what they said.

12 Idioms

1. Keep a look out/ keep an eye out/ keep your eyes open (for something); keep looking; watch out; try to spot something; be aware/show awareness; pay attention; keep focused; concentrate on something.
2. Feeling nervous/worried/ scared/ excited/ funny tummy/feeling sick.

3. I’m not going to say anything/ talk/ tell any one; I won’t say a word; I won’t tell/ give away the secret; I’m being quiet; no speaking.
4. Let’s go/ get going; drive off; go/jump in the car (and drive off); leave somewhere; drive on the road.
5. Stop talking/ be quiet/ shut up/ shush/ don’t speak/ stop talking nonsense/ stop moaning.
6. This is easy.
7. Overjoyed/ very happy/ ecstatic/ delighted/ glad/ excited.
8. He’s having a great/ good/brilliant time/ fun;/ really enjoying it; he’s having a ball/ a boss time/ having the time of his life.
9. She’s grumpy/ in a grump/ cross/ not being very nice/ moody/ narky; she’s got a bad attitude/ out of sorts/ not acting how she usually does; she’s a bit stressed; she’s being a bit mean; she’s not very happy.
10. He’s annoying (me)/ getting on my nerves/ driving me mad/insane/ nuts/ making me cross/angry; he’s frustrating me/ bugging me; driving me bonkers/bananas; winding me up; making me stressed.
11. Hurry up; get a move on; move/go faster/quicker; look sharp; look lively; get moving; get a wiggle on.
12. It’s not my (kind of) thing/ my style; I’m not fond of it/not keen on it/ not really into it/ don’t like it; it’s not my favourite; it doesn’t really suit me; it’s not really for me; this isn’t my kind of work; it’s not my forte;.
13. They get on (really) well; are really good friends/pals; are best friends; are a good couple.
14. He’s going to start a new/better way of doing things; a new way of life; start again/over/afresh/ have a clean start; he’s going to change his personality; he was bad but now he’s going to be good; his attitude is going to improve; he’s going to start acting differently; he’s having a clean slate; moving on; do better; be a better person.
15. Telling a joke/ joking/ teasing/ winding me up/ making it up/ tricking me/ kidding me; having a laugh; telling a bit of a lie.
16. I’m (very) busy; got lots/too much on; got lots of jobs to do; I’m in the middle of doing something; I’ve got my hands full; I’ve got a lot on my plate; I am occupied; I can’t make it at the moment.
17. She will be very angry/ mad/ annoyed/ furious/ fuming; she’ll go crazy/ nutty/ crackers/ beserk.
18. I need time to think about it; I need to consider a problem for longer.
19. He’s not listening/ not paying attention; he forgets (what you have said) straight away; he ignores what you’ve said; you don’t remember something; you’ve not taken notice.
20. What you said is just/exactly/absolutely right/correct/spot on; you’ve got it (just right); you’re bang on; if you’ve done something really well/ done a good job.

Appendix xii

March 2010 **<u>Invitation to Take Part in a Pilot Study</u>**

Study Title- 'The development of a British assessment of inferential comprehension for primary school aged children.'

I would like to invite children in your school to take part in a pilot project related to the larger study named above.

Why is the above study being done?

My experience as a children's Speech and Language Therapist (SLT) has highlighted a gap in this area of language assessment. Some children with language difficulties can understand spoken sentences but are not able to make inferences from information given to them and their knowledge of the world. There is very little research evidence currently available for the typical development of inferential understanding. This study aims to identify typical performance at different age levels in this area.

Why are children aged 4-11 years being asked to take part in this pilot project?

The pilot study needs to assess the whole age range of primary school children to determine what age groups it is most appropriate for. I will see around 10 children from each year group individually for two sessions of about 30 minutes. I will be asking them a series of questions.

Who is in charge of the study?

The study is being carried out by Anne Hewitt. I am an experienced Speech and Language Therapist and have worked with children and adolescents for the past 21 years in Sefton PCT. I also work part-time at Manchester Metropolitan University (MMU) as a senior lecturer. I will be carrying out this study as a PhD project and am being supervised by Professor Jois Stansfield (Speech and Language Therapist) and Professor Juliet Goldbart (Psychologist). The assessments would be carried out at the end of the Spring term and in the Summer term 2010. I anticipate needing to spend around 8 days in total in the infant school. I would need a small room to work in.

If you are interested in the study and are happy for children in your school to take part in the pilot project, please read the attached Pilot Study Information for Parents sheet and contact me on the phone number or e-mail below.

Thank you for your time. Anne Hewitt, M.Ed, Senior Speech & Language Therapist and Senior Lecturer, MMU

Appendix xiii

October 2010

Pilot Study Information for Parents

This study aims to develop a British assessment of inferential comprehension for primary-school children.

I would like to tell you about some research that I am asking your child to take part in. This project has been approved by Manchester Metropolitan University's (MMU) Academic Ethics Committee.

What is the project about?
I am trying to find out about primary school children's understanding of complex spoken language. I will be assessing a few year 3 children in early January 2011.

What will happen if my child takes part in the pilot project?

I will see your child for two sessions of about 30 minutes at school and ask him/her a series of questions. I will write down your child's answers. I will NOT audio- or video-tape your child. When your child is finished, I will thank him/her and give a sticker.

The only information I need about your child is their school year and the month and year of their birth. If you agree to their participation, this information will be provided by school. Your child's full name will NOT be put on the assessment sheets.

What will happen if my child does not take part in the pilot project?

There are no consequences for not allowing your child to participate in this study. However, I do hope you will agree to their inclusion.

Who is in charge of the pilot project?
The study is being carried out by Anne Hewitt. I am a Senior Speech and Language Therapist with Sefton PCT and have worked with children for 21 years. I am also a part-time senior lecturer at MMU. I have a current enhanced CRB check with Sefton PCT.

What happens with the work that my child does?
The results of the pilot project will help me to develop a more accurate assessment of advanced understanding that I will be able to try out on larger numbers of children.

What if I have questions about the pilot project?
If you would like to ask any questions or get more information about the project, I would be happy to speak with you. Please contact me using the details at the bottom of this information sheet.

Thank you for time.

Anne Hewitt, M.Ed., Senior Speech & Language Therapist,
Senior Lecturer in Speech Pathology & Therapy
Manchester Metropolitan University

Address:

Please fill in your child's name and return this slip to your child's teacher ONLY if you would NOT like your child to participate

----
-

Pilot study for the development of a British assessment of inferential comprehension for primary-school children.

I would <u>**NOT**</u> like_______________________(child's name) to participate in the pilot research project
Class ________
Signed __________________________ (parent/guardian)

Appendix xiv

<u>Invitation to Take Part in a Language Study</u> January 2012

Study Title- 'The development of a British assessment of inferential comprehension for primary school aged children.'

I would like to invite a few children in your school to take part in a project related to the study named above.

Why is the above study being done?

My experience as a children's Speech and Language Therapist (SLT) has highlighted a gap in this area of language assessment. Some children with language difficulties can understand spoken sentences but are not able to make inferences from information given to them and their knowledge of the world. There is very little research evidence currently available for the typical development of inferential understanding. This study aims to identify typical performance at different age levels in this area.

What is involved in the study?

I need to assess primary school children aged 5-10 years. I have already completed a pilot study on 50 children but now need to trial the final version on a lot more. I am requesting that you allow a graduate-SLT who has a current CRB check to carry out these assessments in one year group in your school. She will see the children individually for around 30 minutes and will be asking them a series of questions. She will need a quiet space to work in. This does not need to be a separate room. A quiet section of a corridor would be fine. The assessments would be carried out from February to July 2012 on days to suit the graduate and the school. She would be in the school for no more than 10 days during these 6 months.

Who is in charge of the study?

The study is being carried out by Anne Hewitt. I am an experienced Speech and Language Therapist and have worked with children for the past 21 years in Sefton PCT. I also work part-time at Manchester Metropolitan University (MMU) as a senior lecturer. I will be carrying out this study as a PhD project and am being supervised by Professor Jois Stansfield (Speech and Language Therapist) and Professor Juliet Goldbart (Psychologist). I have recent enhanced CRB clearance as do the graduate-SLTs.

If you are happy for children in your school to take part in the project, please read the attached Research Study Information for Parents and let the graduate-SLT who has contacted you know. Thank you very much.

Anne Hewitt, M.Ed, Senior Speech & Language Therapist and Senior Lecturer, MMU

Appendix xv

November 2011

Research Study Information for Parents

This study aims to develop a British assessment of inferential comprehension for primary-school children.

I would like to tell you about some research that I am asking your child to take part in. This project has been approved by Manchester Metropolitan University's (MMU) Academic Ethics Committee.

What is the project about?
I am trying to find out about primary school children's understanding of complex spoken language. I (and some student- or graduate- Speech and Language Therapists) will be assessing as many 5-11 year-old children as possible from each year group in the Autumn, Spring and Summer terms, 2011 – 2012.

What will happen if my child takes part in the project?

I will see your child for one session of about 30 minutes at school and ask him/her a series of questions. I will write down your child's answers. I will audio- tape a small number of children so that my supervisors can check if my scoring system works. I will only do this if you have given specific consent for your child. When your child is finished, I will thank him/her and give a sticker.

The only information I need about your child is their school year and the month and year of their birth. If you agree to their participation, this information will be provided by school. Your child's full name will NOT be put on the assessment sheets. Any audio-recordings of your child will be kept in a locked secure filing cabinet and will be destroyed once the PhD has been completed.

What will happen if my child does not take part in the project?

There are no consequences for not allowing your child to participate in this study and you can withdraw your child at any time. However, I do hope you will agree to their inclusion.

Who is in charge of the project?

The study is being carried out by Anne Hewitt. I am a Senior Speech and Language Therapist (SLT) with Sefton PCT and have worked with children for 21 years. I am also a part-time senior lecturer at MMU. I have a current enhanced CRB check with Sefton PCT. I work in one of the units in the school every week so am already a familiar face to the children in the school. Any students or graduate SLTs who help out with the assessments will all have a current CRB check.

What happens with the work that my child does?

The results of the project will help me to develop an assessment of advanced understanding in primary-aged children.

What if I have questions about the project?

If you would like to ask any questions or get more information about the project, I would be happy to speak with you. Please contact me using the details at the bottom of this information sheet.

Thank you for your time.

Anne Hewitt, M.Ed., Senior Speech & Language Therapist,
Senior Lecturer in Speech Pathology & Therapy
Manchester Metropolitan University

Address:

Please fill in your child's name and return this slip to your child's teacher ONLY if you would NOT like your child to participate in the study.

Research study for the development of a British assessment of inferential comprehension for primary-school children.

I would **NOT** like____________________________(child's name) to participate in the research project

Class _______________

Signed _______________________(parent/guardian) Date _______________

Please fill in your child's name and return this slip to your child's teacher if you give permission for me to audio-tape your child's responses.

Research study for the development of a British assessment of inferential comprehension for primary-school children.

I DO WANT my child ___________________ (child's name) to be included in the research project and I give my permission for his/her responses to be audio-taped.

Class ____________

Signed ____________________________ (parent/guardian) Date _______

Appendix xvi

<u>Procedure for Student and Graduate SLTs to follow for the PhD data collection</u>

1. Decide on which local school you are going to approach. If you have a choice please select a school where the number of free school meals is more than 20%. Contact the Head Teacher or Deputy Head Teacher of this local primary school. Introduce yourself, tell them about the study, say you have a current CRB check and ask if they would be happy for you to carry out the assessment with 1 year group. When you let me know that you have permission from a school to carry out the study I will tell you which age group to target (5:0 up to 9:11). Let me know if you have a preference for a particular year group.

2. If they agree, e-mail them the 'Invitation to take part in a language study' and 'Research information for parents' information sheets. The study is set up as an 'opt out' study (i.e. parents only send the form back in if they DO NOT want their child to take part in the study). If, after reading this information, the school insists that it is made into an 'opt in' study let me know and I will send you a modified version of the consent form so that parents send it back if they DO want their child to take part.
(Of the 5 schools I have approached so far only 1 has insisted on an 'opt in' format). The school will ask for proof that you have a current CRB check. You legally have to only provide your CRB number and the date the check was completed. Schools are not supposed to ask to see a copy of your CRB certificate. If they insist on this and you are happy to show it then do so. The school is not allowed to take a photocopy of your CRB certificate.

3. Once the school is happy with the information ask them to send the 'Research information for parents' sheets out to all the parents in the year group you are covering. The school is usually happy to photocopy these forms but if they are not let me know how many you need and give me the address you want me to send them to and I will send you the required number. Check who the Head teacher wants you to liaise with in the school to carry out the study. It is usually the SENCO as they need to tell you which children have special needs as these will be excluded from the study. The study is establishing norms in typically developing British children so not all children in the school will be included. Ask the Head Teacher at this stage what percentage of children in the school get FREE school dinners (NOT what percentage of children in the school eat hot dinners, as 1 school recently told me!). This is an indicator of the socio-economic banding of the school and is VITAL for me to have to ensure I cover a range of areas.

4. Liaise with the school SENCOs. Tell them about the study and give them a copy of the information letters. Tell them which year group you will be assessing. Ask them to let the class teacher(s) of that year group know about the study, and if you have time go and introduce yourself to that class teacher and explain the study. Ask the SENCO to let you have a class list of all the children in that year with any children with significant special needs

(statemented or having 1:1 support); any receiving Speech and Language Therapy; and any for whom English is not a first language (and they are not yet proficient in English) marked on the list. You will need the children's full names and date of birth on the list.

5. Let me know when you have got to this stage (phone, text or e-mail me) and I will send you out 10 test forms. Give me the address you would like me to send these forms to. If you are not able to carry out 10 assessments just do as many as you can. If you need any more than 10 forms just let me know. I have put in the letter to the schools that you will be in for a maximum of 5 days from June to September 2012. You should be able to carry out 5-6 assessments per day so you are only likely to be in the school for around 2-3 days.

6. Wait about 2 weeks from the time the forms are sent to parents to give them time to send back any 'No' forms.

7. Go into the school and collect all the 'No' forms returned by parents and cross these children off your list. Also cross off any of the children identified by the SENCO as having special needs, SLT or insufficient English.

8. Arrange with the class teacher the days that you will be coming into the school and identify a quiet place where you will be able to take the child. Ask if the children in this year group like stickers and what type. Purchase some of these stickers (I will reimburse you for these).

9. Go into school on the designated days. Check just beforehand that the class isn't out on a trip or involved in a special activity that day. Select a cross section of ages of children from the class from the youngest to the oldest (e.g Year 1: 5:10, 5:11, 6:00, 6:01, 6:02, 6:03, 6:04 etc. up until 6:11) and alternate the sexes (boy-girl-boy-girl). Aim to get 1 boy and 1 girl of the same age.

10. Select your child. Check with the teacher it is OK to take that child. Explain to the child that you are just going to ask them a few questions, it will take 30-40 minutes and they will get a sticker at the end if they want one. Use the following script with the child:
"Hi, my name is ----- and I am a student/graduate Speech and Language Therapist. Do you know what that job is? (Usually the child says No). The job is to help children who find it hard to understand and talk. I know you don't have any difficulties with talking. This test has been made for children who have problems with talking but first it has to be tried out with children who don't have any problems, like you. Are you happy to help me with this?"
If they are not then ask another child (I haven't had any child refuse yet).

11. Take the child to the quiet place and begin the assessment. Explain that you will be writing down their answers so it will take a bit longer.

12. Follow the guidelines on the test form. Don't forget to put the child's month and year of birth and exact age in years and months on the form and include the percentage of free school dinners received in the school.

13. The test should take around 35-40 minutes to administer but it may take longer than that the first few times you do it. Older children will be able to manage to do it all in one session but you may have to see younger children (particularly reception children) in 2 shorter sessions with a break in between.

14. Complete the test. **Please write in your clearest handwriting as close to printing as you can make it**. Thank the child and give them a sticker if they want one and go and get the next child.

15. Repeat until you have completed all the tests you are able to do.

16. Ring me on my mobile or e-mail me if you have any queries as you are carrying out the assessments.

17. Score up **only one** of the assessments you have done using the marking criteria sent to you. Write any comments about any difficulties you had with the scoring on the 'any comments' sheet sent to you.

18. Put all the completed assessment forms in a securely sealed envelope and send them to my address (given below). Ask for a 'proof of postage' receipt. I will reimburse you the postage cost.

19. Enclose your travel claim form, postage receipt and receipt for any stickers purchased and I will process these at MMU and you will (eventually) be reimbursed for these costs. Also include any comments you have on the assessment (the 'any comments' sheet is an e-mail attachment).

20. Please let me know at any stage if you are unable to continue carrying out these assessments.

Thank you so much for helping me out with this data collection. I will acknowledge you in my write up. Don't forget to add this to your CV.

Address to send the completed forms back to:
Anne Hewitt, Speech Pathology Lecturer, Health Professions Department,
Or you can return them to me at MMU at the beginning of next term.

Appendix xvii

Immature or unusual responses to test questions

Colour coding
Red = 5:00-5:11
Blue = 6:00-6:11
Green = 7:00-7:11
Purple = 8:00-8:11
Brown = 9:00-9:11

1) Simple Deductions

1. I am a musical instrument. You bang me with your hand or with sticks. I make a loud noise.
   What am I? (drum/ tambourine/ cymbal/ gong)
2. I am an animal. You would see me in a zoo. I look like a horse. I have black and white stripes.
   What am I? (zebra)
- Horse (girl 5:03, girl 5:10)
- Goat (girl 5:10)
- A donkey (boy 6:08, boy 7:09)
- A tiger (boy 6:09, girl 7:03, boy 8:11, girl 9:00)
- A lion (boy 7:07)
- A cow (girl 9:11)

3. I am an animal. I am pink or brown. I live on a farm in a sty. People say I am greedy. I have a curly tail.
   What am I? (pig)
- A fox (boy 5:01)
- A chicken (boy 5:02)
- A horse (boy 5:08, girl 8:05, boy 9:05)
- A giraffe (girl 5:08)
- A monkey (girl 5:10)
- A flamingo? Like a little bird (boy 6:01)
- A cow (boy 6:05)
- A sheep (boy 7:05, boy 7:09)
- A dog (boy 9:11)

4. I am clothes. I am often made from wool. You put me on your hands. I keep your hands warm when it is cold.
   What am I? (gloves/ mittens/ muff)
- A squirrel (boy 5:01)

- Jumper (girl 5:03)
- A sheep (girl 5:04, boy 5:11, girl 5:11, boy 6:00, boy 6:01) Multiple X
- An egg (boy 5:08)
- An owl (boy 5:09)
- A bird (girl 5:10)
- A hen (girl 5:10)
- The fluffy bear (girl 5:10)
- A coat (boy 6:09)
- A woollen jacket (girl 9:05)

5. I am food. I have a hard shell. You sometimes need to crack my shell to get at me. There are lots of different kinds of me. Squirrels like to eat me.

What am I? (nut, any kind)

- A bird (girl 5:04)
- Eggs (girl 5:08, girl 5:10) Multiple examples
- A snail (2 x girls 5:10, girl 6:08, boy 7:04, girl 7:10, girl 8:07)
- A turtle (2 x boys 6:01) A tortoise (boy 7:11)
- A pine cone (boy 6:05, girl 6:06, girl 6:08, boy 7:07)
- A lobster (girl 6:08)
- A crab (girl 6:09)

6. You find me inside and outside buildings. I have a frame. You can see through me. If you throw a hard ball at me I could smash.

What am I? (window)

- Spider (boy 5:01)
- A robot (boy 5:02)
- A cow (boy 5:03)
- A crocodile (boy 5:08)
- Mirror (girl 5:10 and many other 5 year-olds) Multiple times
- A tower (boy 6:00)
- A picture (boy 6:01)
- A Christmas tree (boy 6:02)
- A door (boy 6:05, boy 7:05)
- Glass (3 X boys 6:06) Multiple times
- Bus stop (Girl 6:02)
- A vase (boy 7:01, boy 9:08)
- A snail (boy 7:05)
- A television (boy 7:08)
- A picture frame (boy 8:00, girl 9:02, girl 9:08) Photograph (girl 8:04)
- A car (boy 8:02)
- A tennis ball (boy 8:07)

7. I am a musical instrument. I am made of wood. I have strings. You play me with a bow. You hold me under your chin.

What am I? (violin/ viola/ fiddle)

- A guitar (girl 5:04, boy 6:01) Multiple times
- A drum (girl 5:10)
- A scarf (boy 6:00)
- A trumpet (boy 6:02)
- Like a guitar with a little stick (boy 6:06)
- A harp (boy 8:05)
- A violet (boy 8:07)
- A lute (boy 9:10) ?

8. I am made of metal and waterproof material. I have a handle. You put me up and down. I keep you dry in the rain.

What am I? (umbrella/ brolly)

- A cloud (boy 5:03)
- A shelter (girl 5:04, boy 6:05)
- A door (girl 5:10, boy 6:01, boy 8:04)
- A roof (girl 5:10)
- A door handle (boy 6:00)
- A hat (boy 6:06)
- A rain coat (boy 6:08) A waterproof coat (girl 6:08, boy 8:07)) A coat (boy 6:09) Multiple times
- A door (boy 6:08, boy 7:09, girl 7:00)
- A watering can (girl 6:01) (distracted by a child running past)
- A teapot (girl 6:02)
- A hood (boy 7:05, girl 7:05, girl 8:00, girl 8:05, boy 9:05, boy 9:06) Multiple
- A zip (girl 7:06, girl 8:10)
- A car (girl 7:06, boy 8:11)
- A window (boy 8:04)
- A tent (boy 9:05)
- It sounds like a bicycle or car (boy, 9:06)

<u>2) Why Question Inferences</u>

1. Why is it silly to hang washing out to dry when it is raining?

- Cos you need an umbrella (boy 5:05)

2. Why do you need a towel when you go swimming?

- Cos everyone'll see your bum (prompt -> drying) (boy 7:04)
- Cos if you didn't it'd be rude cos everyone would see (girl 8:05)

3. Why do people have jobs?

- That's why so the fings won't fall over and they won't get boken (boy 5:00)
- Because they are busy (boy 5:03, girl 5:03)

- So they get very clever (girl 5:07)
- Because or they won't get to be parents (Boy 6:00)
- They need to clean up (girl 6:04)
- So they can have driving lessons (Boy 6:09)
- So they have something to do all day (girl 8:06)

4. Why mustn't you play with matches?

- Cos it's really important (girl 5:03)
- Cos you got to match them up again (boy 5:05)
- Cos they're boyish (girl 5:10)
- Cos you get hurt and they're quite heavy (boy 8:04)

5. Why are mirrors made out of shiny surfaces not cardboard?

- Because if it was too windy the cardboard would blow off it (boy 5:03)
- Cos you can't see how you feel well and you're looking good (boy 5:05)

6. Why can't you read in the dark?
7. Why are windows made out of glass not bricks?

- Because they'll blow your house down (girl 5:03)
- Cos you got to see in the mirror to see if you're looking cool (boy 5:05)
- You could bang your head on the bricks (girl 8:07)
- Because it would be too hard to open (boy 9:08)

8. Why is a pan made out of metal not chocolate?

- Cos you can't make chocolate in the pan because it doesn't need cooked (girl 5:03)
- Because if you ate it you won't be able to draw on it (boy 5:07) (thought 'pen'?)

9. Why are wellies made out of rubber or plastic not paper?

- Cos if you took them off in the wind they'd blow away (boy 5:03)
- Cos you can't stand on it (girl 5:03)
- Cos they will hurt you (girl 5:10)

10. Why shouldn't you agree to take a lift in a car from people you don't know?

- Cos if you put a car in a lift the car would break while the doors go in on it and they'd go on fire (boy 5:04)
- Cos you have to get a taxi (boy 5:05)
- You're in the rain and if you haven't got hood you'll be really wet (girl 5:09)
- Because they might stranger-danger you (boy 5:10)

11. Why are teeth covered in enamel not cotton wool?

- Because they've got to move and you've got to go in the farm to walk properly (boy 5:01)
- Because cotton makes you sneeze and they don't (girl 6:01)
- Cos you can wash enamel off and not cotton wool (boy 8:04)

12. Why do we need pedestrian crossings on roads?

- The cars stop and let the zebra past if it has escaped from the zoo (girl 6:09)
- So it would show up in the dark and the ladies can stand on it to show you where it is (boy 8:04)

- If you walked across the road you might not see if no zebra's crossing (girl 8:05)

3) Making Predictions Inferences

1. What would happen if you left a block of ice in the sun?
- The world will be fire (boy 5:00)
- You would be blind cos the sun would get in your eyes (boy 5:11) (?heard ice as eyes?)

2. What would happen if you were very late for school every day?
3. What would you do if your friend got stuck in some mud?
4. Sam and Ruby had a date to go to the cinema but Ruby didn't turn up. How did Sam feel?
5. What would happen if you took away the bottom tin in a stack of tins at the supermarket?
- Because you'll get fat (girl 5:03)
- You won't have anything to eat (girl 5:04)
- It'll be stealing (girl 5:04)
- The pan would feel lonely (girl 5:05)
- Happy (boy 6:04)
- The supermarket people will ring the police (boy 6:06)
- You'd get arrested (girl 6:11, boy 7:05)
- You'd be stealing, you'd be a thief (boy 7:01)
- You'll get caught by the police (girl 8:04)
- If you thirsty (boy 8:05)

6. What would happen if you had a glass of water full up to the top and then you added 6 ice cubes?
- You'll get really hot and hot and hot (girl 5:03)
- It'll displode (girl 5:04)
- You a cold brain (girl 5:10)
- Happy (boy 6:04)
- You'll feel sick (boy 6:06)
-

4. Formulating Solutions Inferences (What could you do?)

1. What could you do when you are cold?
- Freeze (boy 8:07)

2. What could you do if you wanted to buy a toy but you didn't have enough money?
- Just count on from 10 to 11 (boy 5:10)
- I would go and say to the witch people 'Please may I have some money' (girl, 6:08)

3. What could you do when you are hungry?
4. What could you do if you wanted to eat your soup but it was too hot?
- Go on a trampoline for 5 minutes (boy 8:07)

5. How could you get to the top floor of a very tall block of flats?

- Cos if you go right to the top you fall and bang your head (boy 5:01)
- Wiv a ladder (boy 5:03)
- You can't get back up there cos you haven't got no ladders (girl 5:03)
- You'd get taller and you'd climb up your bed and it's no time in your PJs (boy 5:06)
- You can go up in a helicopter and fly up (6:06)
- Run (boy 7:05)
- Strength (boy 8:05)

6. What could you do if you dropped a glass of milk all over your kitchen floor?

- You would sip (slip) over (boy 5:02)
- You will flood. There'll be a flood (boy 5:11)

7. What could you do if you burnt a cake you had baked for your mum's birthday?

- Get the hose pipe and just put it out or if you had a fire extinguisher you could put it out (boy 6:04)
- Pretend it's a chocolate cake (boy 7:09 (prompt -> Make something else like cupcakes maybe - correct)

8. How do we keep milk fresh and cold?

- Wiv the dryer (boy 5:03)
- Putting a lid on (girl 5:04)

9. What could you do if you accidentally broke something that you had borrowed from a friend?

- You would hide it (boy 5:03)
- Nothink. You'll be in big trouble (boy 5:11)
- Keep quiet (girl 8:03)

10. How could you talk to somebody who lives hundreds of miles away from you?

- Drive to their house (girl 5:04)
- You could copy their language – what they're saying (girl 5:05)
- Go in an aeroplane (girl 5:10)
- Run all the way or on a bike (boy 6:08)
- Get a car (boy 7:06)

11. How could you find out what your teacher looked like when s/he was a little girl/boy?

- You look fat (girl 5:03)
- Find a machine that goes back in time (boy 6:06) Go back in time (boy 8:07)
- You shouldn't as her – you should always think (girl 6:05)
- Little (boy 7:04)
- Contact the school (girl (9:02)

12. What could you do if you locked yourself out from your house and nobody else was at home?

5. <u>Explaining Inferences (How do you know?)</u>

1. How do you know that someone is tired?
- Cos if their eyes are ugly (boy 5:07)
- You see these black things and they fall asleep (boy or girl 6-6:11)
- Their eyes go drifty (girl 7:07)
- They're all lazy (boy 9:08)

2. How do you know that a group of children want you to come and join in their game?
3. How do you know that someone is angry?
4. How do you know that someone has been eating biscuits in the kitchen?
- You gotta go tell your dad and then they gotta get their coats on and go away cos they smashed something (boy 5:05)
- Cos it might be a mess (boy 5:07)
- They'll be fat (boy 7:11)

5. How do you know that someone is too hot?
6. How do you know that someone has put too much bubble bath in the bath?
- It'll dust expode (boy 5:00)

7. How do you know that your neighbours are moving house?
- Their house is going to break down (boy 5:03)
- Because you can see them go in the caravan (girl 5:10)

8. How do you know that someone has got toothache?
- Cos they get all muddy (boy 5:01)
- You'd have to get back in for your mum because your answer is 'No' (girl 5:03)
- Because you can't hear them properly (girl 7:06)

9. How do you know what's inside a box of something at the supermarket?
- Cos the supermarket's not open (girl 5:03)
- Use your senses. You could see a tail if it's a cat. You could like …. I don't know what else (girl 7:08)
- It would be moving (boy 8:00)

10. How do you know that someone doesn't want you to hear what they are saying to someone else?
- They're bored (boy 5:03)
- You can't have an /atitud/ (attitude) (girl 5:05)

11. How do you know that someone thinks there is a bad smell in the room?
- Someone had a flat water to add to a smelly room (boy 5:01)
- It might be poo (boy 5:03)
- Just get police to smell it and then put water on them (boy 5:08)
- It stinks – like I trumped in school before (Boy 5:11)
- Tos someone's farted (boy 6:06)

12. How do you know that someone thinks you are playing your music too loudly?

6) Making Inferences from short passages

1. Dad hung the washing out on the line but when he went to bring it back in it wasn't there.
What had happened?
2. Raj was playing on the beach. He trod on something and had to go to hospital?
What did Raj tread on?
- A conker (boy 7:07)
- A hedgehog (boy 8:00)

3. Ellie paid the money to get in. She got changed and put her clothes in a locker. She put the key band on her wrist and took her towel with her.
What was Ellie doing?
- She was trying to get big (boy 5:01)
- She was trying to get her clothes and someone wobbed them off her (girl 5:03)
- She was fixing the man (boy 5:05)
- She was trying to get in (prompt) the ambulance (girl 5:10)
- Stopping people getting in (Boy 6:04)
- Maybe they had a pet and the pet escaped (girl 8:07)

4. Alisha and her mum went to see a concert but had to come back home without seeing it.
What happened?
- Her dog didn't come wiv her (girl 5:01)
- Someone come and wobbed the cookies and all the food so they can't get no cash (girl 5:03)

5. Dan's dad took him to somewhere outside. He pushed Dan up and down on something and pushed him round and round on something else. Dan climbed up some steps and went down something.
Where were Dan and his dad?
- At the park or at wugby (boy 5:00)
- In the doctor's (boy 5:05)
- The attic (boy 5:08)
- On the lift (girl 5:10)
- In the basement (8:00)

6. A group of friends went out to a party together. They had planned to get the bus home afterwards but ended up taking a taxi.
Why?
- Cos the bus wasn't coming because it had too much stuff in (girl 5:06)
- Buses are tall, taxis are small (boy 8:05)

7. Jane and her friend Sasha went out for a bike ride. They came back later than they were supposed to and Jane was pushing her bike.
Why?
- Cos she didn't know how to wide (ride) without stabilisers (boy 5:03)
- Cos she couldn't ride cos she was a baby (girl 5:04)
- Cos she was sleeping (boy 5:08)

- They pushed somebody off (boy 8:05)

8. James and his dad are sitting in the front row. The curtain goes up and people in costumes come on and talk loudly and sing.
Where are James and his dad?

- They're in the beach and the park and they fell off and Izzy came over to say 'What've you been doing?'(girl 5:03)
- Cos they were on Halloween (boy 5:10)
- On the bus (girl 6:02)

9. Mum and dad have just eaten a nice meal in an expensive restaurant. They ask for the bill and then both look worried and embarrassed when it comes.
Why?

- Because they're so young (girl 5:04)
- In case grandma and grandad are looking after the children fought that they could have afters. Actually, I don't know (girl 5:05)
- Cos they might have done something the police wanted them to do (boy 5:08)

10. Dad bought a chicken to cook for Sunday dinner but the family ended up eating just vegetables and gravy.
What happened?

- Chicken ran away (boy 5:01)
- They won't get more energy (girl 5:03)
- They were all chubby (boy 5:05)
- I bought a real chicken from a farm and it never went in the oven and it ran away (boy 7:06)
- It could be Monday (girl 7:08)

7) Situation-based emotions inferences

1. Rachel's pet cat died. How does she feel? (sad)
2. A boy took a toy off Jake. How does Jake feel? (angry)
Happy (boy 5:08)
3. A big fierce dog chased Raj down the road. How does he feel? (scared)
4. Lucy's dad bought her some sweets. How does she feel? (happy)
5. Alice was invited to a friend's birthday party. How does she feel? (happy)
Happy (girl 5:00)
6. Jamelia painted a lovely picture but a girl came and scribbled on it on purpose. How does Jamelia feel? (angry)
7. Stefan broke his favourite toy. How does he feel? (sad)
8. Three year-old Tom has lost his mum in a busy shopping centre. How does he feel? (scared)
9. A boy hits Gemma's sister at school. How does Gemma feel? (angry)
Happy (boy 5:04)
Happy if she hates her sister or angry (boy 9:06)

10. Jamie's dad takes him for tea at McDonalds but McDonalds is closed. How does Jamie feel? (sad)
Scared (girl 5:10)
11. Rahal's friend has brought him a present back from his holiday. How does Rahal feel? (happy)
Sad (girl 5:00)
Angry (girl 5:04)
12. Katy is alone in a dark, empty house. How does she feel? (scared)
13. Laura got told off by the teacher but it was another child's fault. How does Laura feel? (angry)
Happy (girl 5:07)
14. Jake has won first prize in a competition. How does he feel? (happy)
Happy (girl 5:00)
15. Sophie is stuck in a tree, about to fall. How does she feel? (scared)
16. Shazia's family holiday is cancelled at the last minute. How does she feel? (sad)
Happy (boy 5:09)

8) Belief-based emotions inferences

Matthew wants fish and chips for tea and he will be getting fish and chips.
Matthew thinks there are fish and chips for tea.
1) How does he feel? (happy).
2) His mum gives him fish and chips for tea. How does he feel? (happy)

Raj wants a toy plane for his birthday but he is getting a teddy.
Raj thinks he is getting a toy plane for his birthday.
3) How does he feel? (happy).
4) He gets a teddy for his birthday. How does he feel? (sad)
Jane wants to go to the beach but she is going to the library.
Jane thinks her dad is taking her to the library.
5) How does she feel? (sad)
6) Her dad takes her to the library. How does she feel? (sad)

Samia wants to go clothes shopping with her mum and she will be going clothes shopping.
Samia thinks she is going to help her mum clean the house.
7) How does she feel? (sad)

- Happy that she's going, but not happy that she thinks she will be helping her mum do the washing – was it washing? (boy 7:10)

8) Mum takes her clothes shopping. How does she feel? (happy)

9. Mental state verb inferences

1. Karen knows that the door is locked. Is the door locked? (yes)
2. Jack thinks that the door is locked. Is the door locked? (maybe)
3. Raj manages to lock the door. Is the door locked? (yes)
4. Shazia forgets to lock the door. Is the door locked? (no)
5. James wants to lock the door. Is the door locked (maybe)
6. Jasmine realises that she locked the door. Is the door locked (yes)
7. Zak is sorry that he locked the door. Is the door locked? (yes)
8. Alisha promises to lock the door. Is the door locked? (maybe)
9. Dan remembers to lock the door. Is the door locked? (yes)
10. Jamal learns that the door is locked. Is the door locked? (yes)
11. Ellie believes she has locked the door. Is the door locked? (maybe)
12. Billy is happy that he locked the door. Is the door locked? (yes)

10) Strange Stories

Pretend

John and Rashid are playing spacemen. John picks up a bowl and puts it on his head. He says: 'This bowl is a space helmet.'

1) Is it true what John says?
2) Why does he say it?

Because he's imaginating (girl, 7:02)
Because they don't wear that otherwise they'd die – no oxygen (boy 8:11)

Lying

Julie hates going to the doctor. She has very bad stomach pains and is holding her stomach. Her mum asks "Have you got tummy ache?" Julie says 'No'.

3) Is it true what Julie says?
4) Why does she say it?

- So she can carry on going to the shops (boy 8:05)
- Cos she might be pregnant (boy 9:05)

Misunderstanding

Jack hurt another child on purpose at school, but no teachers saw him do it. At the end of the day he started to walk home. A teacher ran after him shouting 'Stop Jack!' Jack turned round and said 'I'm really sorry for doing it Miss'. The teacher only wanted to give him his lunch box that he had forgotten.

5) Why did Jack say 'I'm really sorry for doing it Miss'

That's why he doesn't want to go on wed (red) (Boy 5:00)

6) Did the teacher want to tell Jack off?

7) How did the teacher feel when Jack said ‘I’m really sorry for doing it Miss’

- Scared (girl 8:10)

(If the child gives the wrong answer to the above say: ‘I think she was surprised.’ Then ask ‘Why?’)

8) Why?

- Because the teacher never knew who done it and he just said it out off of nowhere and it came in randomly (boy 9:05)

Contrary Emotions

James’ group of friends is going swimming on Saturday morning. James really wants to go with them but he knows that Sam, a big boy who bullies him at school goes swimming in the same pool on Saturday mornings. When his friends ask him if he wants to go swimming with them he says ‘No.’

9) Is it true what James says?
10) Why does he say he doesn’t want to go swimming?

White lie

Tom’s friend Samia has just had her hair cut. Tom doesn’t like it. He doesn’t think it suits her. He liked her old style better. When Samia asks Tom ‘Do you like my new hair style? Tom says ‘Yes, it’s very nice.’

11) Is it true what Tom said?
12) Why did he say it?
Tos if he said ‘No’ they would like punch him (boy 6:06)
So then she’ll think he likes the gift (boy 8:11)

11) Faux pas

Disco

It was the school disco. Jill was in one of the toilets. Karen and Sara came in to the toilets afterwards. Karen said: ‘Did you see how Jill was dancing. Wasn’t she terrible!’ Then Jill came out of the toilet.

1) How does Karen feel?
2) How does Jill feel?
3) Did Karen know that Jill was in one of the toilets?

Taxi driver

Mike and Rashid are talking about different jobs. Rashid says ‘I hate taxi drivers. They’re all so common and rude.’ Mike says ‘My dad’s a taxi driver.’
4) How does Rashid feel?

5) Did Rashid know that Mike's dad was a taxi driver?
6) How does Mike feel?

- Really sad. But people have different opinions (girl 7:03)

Clothes shop

Laura worked as a clothes shop assistant. A lady with a fat tummy came in and tried on a dress. Laura said 'That looks lovely and you will still be able to wear it when you have had the baby.' The lady said 'I'm not pregnant.'

7) Did Laura know that the lady was not pregnant?
8) How does Laura feel?
9) How does the lady feel?

Vase

Mrs Patel has Mrs Brown, an old friend, over for tea. She hasn't seen Mrs Brown for years. Mrs Brown accidentally knocks over a vase and breaks it. Mrs Patel says 'It doesn't matter. It was a wedding present that I never liked anyway. I can't remember who bought it for us." Mrs Brown says 'I bought it for you.'

10) How does Mrs Brown feel?
11) Did Mrs Patel remember that Mrs Brown had bought her the vase?
12) How does Mrs Patel feel?

- She might be just eatin her dinner or laughing (Boy 6:00)

12) Idioms

1. Keep your eyes peeled
   - It means that there's rats or monsters around (boy 9:05)
2. I've got butterflies in my tummy
   - When you eat lots of sweets (boy 6:09)
3. My lips are sealed
   - Like they're chapped and stuck together (boy 8:04)

4. Let's hit the road
   - Let's det the party started (boy 6:06)
5. Put a sock in it
   - It'll be naughty. That's why. What happened. Somebody dies and it goes boom and the car – it's' fell over (Boy 5:00)
   - You have to dig it up (boy 6-6:11)

6. This is a piece of cake

7. She's over the moon

- A cow jumped over the moon (girl 6:06)
- She's on fire – answers all the questions right (girl 7:06)
- She's really clever (boy 7:08)
- She's really far away (girl 9:06)

8. He's having a whale of a time

- She wants to go away for a little bit (girl 6:02)

9. She got out of the wrong side of bed today

- Maybe she was too early for school and no one was there (boy 6-6:11)
- She's a lunatic (boy 9:06)

10. He's driving me round the bend

11. Get your skates on

12. It's not my cup of tea

- It's not your teabag (boy 7:05)
- I don't really like it – I prefer black coffee (girl 9:06)

13. They get on like a house on fire

(Many children say the opposite)

14. He's going to turn over a new leaf

- He's going to have a baby. He might be married (girl 7:03)
- Gonna get married (boy 8:05)

15. You're pulling my leg

- You're holding me back (girl 9:09)

16. I'm tied up at the moment

(Many children saying 'tired')

- Someone ties them up in prison after getting in a matrix van with other people (girl 6:05)

17. She'll hit the roof

- You can walk in it when you're doing house work (?means building site work?) (boy 6:11)
- Go really fast at work (boy or girl 7-7:11)
- She's big (girl 8:09)
- She's jumping on the trampoline in the house (girl 9:05)

18. I need to sleep on it

19. It goes in one ear and out the other

- Earring (girl 6:02, girl 7:09)

- A piece of string (boy 6:06)
- A machine (girl 6:06)
- A phone (boy 6:09)
- She's magic (boy 7:01) Magic (girl 7:07)
- It means it goes through that one and out that one, like ear wax (boy 7:09)
- One kind is funny and one kind isn't. That was just a guess (boy 7:10)
- I'm not really sure. It's something goes round people (girl 8:02)

20. You've hit the nail on the head

- She's saying you're grounded (boy 7:01)
- You've made me shocked (girl 7:06)
- You've give me a headache (boy 7:09)

Appendix xviii

Exploratory case studies: Additional Details

Child A: HFA. Age 5:06.

HICIT Sub-test scores:

1. Simple deductions (5/8)
2. Why questions (0/12)
3. Making predictions (0/6)
4. Formulating solutions (3/12)
5. Explaining inferences (2/12)
6. Making inferences from short passage (4/10)
7. Situation based emotions (9/16)
8. Belief based emotions (3/8)

HICIT examples from a typical 5½ year old and Child A

1) Simple deductions

1. I am a musical instrument. You bang me with your hand or with sticks. I make loud noise. What am I?

Child with Autism (ASD): A lady.

Typical Child (TC): A drum.

2) Why Question inferences

1. Why is it silly to hang washing out to dry when it's raining?

TC: 'Cos it'll get wet.

ASD: 'Cos had grown small because he is watching the rain – because he is still getting wet.

4. Why mustn't you play with matches?

TC: Because they can burn you.

ASD: Because they have to match together.

6. Why can't you read in the dark?

TC: 'Cos you can't see the words.

ASD: 'Cos I need to turn the light on. Because I need to read higher and that's what happens in my number called 25.

10. Why shouldn't you agree to take a lift in a car from people you don't know?

TC: Don't know; 'cos you don't know the people; 'cos they might be mean; 'cos they might stranger-danger you.

ASD: Because the van has 2 floors called 1 and 2 and the car has 1 floor called 1.

12. Why do we need pedestrian crossings on roads? (Explained like a zebra crossing)

TC: Don't know; so you don't go off the road.

ASD: The zebra crossings are roads that are for zebras to cross – to lead the zebras back to the zoo.

3) Making predictions inferences

3. What would you do if your friend got stuck in some mud?
TC: Help them; tell a teacher/policeman; don't know.
ASD: Because my series 1, 2, 3 and 4 Moshlings have to pull it out.
4. Sam and Ruby had a date to go to the cinema but Ruby didn't turn up. How did Sam feel?
TC: Upset/sad
ASD: Great.
6. What would happen if you had a glass of water full up to the top and then you added 6 ice cubes?
TC: It'll spill over; it'll melt; don't know
ASD: Because the 6 water puddles have to get water to buy them 1,2,3,4 and 5 series Moshlings.

4) Formulating solutions inferences (What would you do?)

4. What would you do if you wanted to eat your soup but it was too hot?
TC: Blow on it; wait for it to cool down.
ASD: Because the food called soup is orange and it's an orange thing that looks like tomato soup.
5. How could you get to the top floor of a very tall block of flats?
TC: Climb/ Wiv a ladder/ Don't know.
ASD: The top floor is 9 and it's the top – the highest. And 1 is the bottom floor. Because 9 is a long way up. (Prompt question: Would you go up the stairs?) ASD: No stairs are too funky ... In a 1 to 11 lift.
8. How do we keep milk fresh and cold?
TC: Put it in the fridge
ASD: Because we need it warm not hot.
10. How could you talk to somebody who lives hundreds of miles away?
TC: Drive(x2)/ Use the car/ Walk into their house/ Don't know/ Phone her
ASD: Because it's too far.
11. How could you find out what your teacher looked like when she was a little girl?
TC: Don't know
ASD: She looked like a bumper girl.

5 Explaining inferences (How do you know?)

1. How do you know that someone is tired?
TC: They yawn
ASD: Because it's getting dark.
2. How do you know that a group of children want you to come and join in their game?
TC: Don't know; they ask you
ASD: Because I have to play with them.
4. How do you know that someone has been eating biscuits in the kitchen?
TC: There are crumbs; Don't know.

ASD: Because the number of biscuits called zero is staying behind in the kitchen. (marked correct)
7. How do you know that your neighbours are moving house?
TC: Don't know; you'll see them; You'll see a sign/van
ASD: Because there's all the houses from 0 to 200 and from 201 to 900 and keep changing people.
12. How do you know that someone thinks you are playing your music too loudly?
TC: Because they can hear it; it's too loud; don't know.
ASD: Because any loony tune or any Moshling might have to help me to do what the tunes are to play it even quietly to stop that.

6. Making inferences from short passages

2. Raj was playing on the beach. He trod on something and had to go to hospital. What did Raj tread on?
TC: A stick; a nail; a stone; a shell; a spike; a ball (x).
ASD: Shell – because the word called shell is a sharp word isn't it? (marked correct)
10. Dad bought a chicken to cook for Sunday dinner but the family ended up eating just vegetables and gravy. What happened?
TC: The chicken burnt (correct); Don't know; they couldn't eat it; the chicken died; they got poorly.
ASD: That's what happens when they get them.

Child B: HFA. Age 6:03

HICIT Total score: 31

Mean score for 6:00-6:11 year-olds is 88

His total score for sub-sections 1-8 is **24**
Total mean score for 5:00-5:11 (out of a maximum score of 84) is **48.3**
Some examples of his unusual responses to questions are given below:

1) Simple Deductions **Score 1/8**

6. You find me inside and outside buildings. I have a frame. You can see through me. If you throw a hard ball at me I could smash.
What am I? Naughty

2) Why Question Inferences **Score 3/12**

10. Why shouldn't you agree to take a lift in a car from people you don't know?
People can't have triangle wheels or rectangle wheels and you won't drive it. Or squares. You need circly wheels to drive it so it can move properly.

3) Making Predictions Inferences **Score 1/6**

1. What would happen if you left a block of ice in the sun?
The sun will be angry and then it will shine someone.

4. Formulating Solutions Inferences (What could you do?) **Score 4/12**

5. How could you get to the top floor of a very tall block of flats?

But you can't get everything balanced on top. You can't go to floor 10 because there's no floor 10 in this world.

5. Explaining Inferences (How do you know?) **Score (1/12)**

2. How do you know that a group of children want you to come and join in their game?

But any people doesn't want any games. But the boy doesn't want to play the game but the rest of the people want to.

6) Making Inferences from short passages **Score (2/10)**

1. Dad hung the washing out on the line but when he went to bring it back in it wasn't there.

What had happened?

Everyone must be gone from the trip, but it was tea time. But the children is a long time there so he went tea in some restaurant.

3. Ellie paid the money to get in. She got changed and put her clothes in a locker. She put the key band on her wrist and took her towel with her.

What was Ellie doing?

Ellie is a good girl but he can't walk hisself but he is 4 years-old. When you are a 4 year girl you can't go by yourself. You need an adult.

7) Situation-based emotions inferences **Score (7 /16)**

8. Three year-old Tom has lost his mum in a busy shopping centre. How does he feel? (scared)

Happy

14. Jake has won first prize in a competition. How does he feel? (happy)

Scared

10) Strange Stories **Score (3 /12)**

Pretend

John and Rashid are playing spacemen. John picks up a bowl and puts it on his head. He says: 'This bowl is a space helmet.'

1) Is it true what John says? No

2) Why does he say it? Because if you be upside down, it will fall.

White lie

Tom's friend Samia has just had her hair cut. Tom doesn't like it. He doesn't think it suits her. He liked her old style better. When Samia asks Tom 'Do you like my new hair style? Tom says 'Yes, it's very nice.'

11) Is it true what Tom said? Yes.

12) Why did he say it? Because the other boy doesn't like hair cutting. He likes to grow it back all day.

11) Faux pas   **Score (1 /12)**

Disco

It was the school disco. Jill was in one of the toilets. Karen and Sara came in to the toilets afterwards. Karen said: 'Did you see how Jill was dancing. Wasn't she terrible!' Then Jill came out of the toilet.

1) How does Karen feel? Sad because she missed the thing out of the dance.

2) How does Jill feel? So happy

3) Did Karen know that Jill was in one of the toilets? He was human, but he had to go to the toilet very quick.

12) Idioms   **Score (0 /20)**

1. Keep your eyes peeled

It means you can't put them back in because you would be died and broken.

4. Let's hit the road

That means everything will be damaged and you can't go back in earth again, but you can make a new earth and astronauts can have for them.

7. She's over the moon

She is over the moon but that's because it is all icy and there is cheese in it.

Child C: Pragmatic difficulties. Age 10.02.

1) Simple Deductions

3. I am an animal. I am pink or brown. I live on a farm in a sty. People say I am greedy. I have a curly tail.

What am I? (pig)

Um - is it a ostrich?

2) Why Question Inferences

5. Why are mirrors made out of shiny surfaces not cardboard?

Um – because those are made out of glasses – you know like glass cups.

6. Why can't you read in the dark?

Because you need to go asleep for school.

7. Why are windows made out of glass not bricks?

Um -

8. Why is a pan made out of metal not chocolate?

So you cook a pancakes.

9. Why are wellies made out of rubber or plastic not paper?

Cos they need to change all my underwear. 2X Because if you um done an accident you need to get change and tell your mummy and daddy or auntie and uncle.

3) Making Predictions Inferences

5. What would happen if you took away the bottom tin in a stack of tins at the supermarket?

Er – they um you might get shouted ….. then the police might come and take you to jail.

6. What would happen if you had a glass of water full up to the top and then you added 6 ice cubes?

Um – I fink if you have – ur – you know – put the cup your know right here and you playing your video game and you put it right here then it will spill and your mummy daddy have to clean up.

5. Explaining Inferences (How do you know?)

3. How do you know that someone is angry?

Um you say I'm sorry (prompt from me) – Step away from them

4. How do you know that someone has been eating biscuits in the kitchen?

Ur – you buy more

5. How do you know that someone is too hot?

Um – you have a nice cold drink (prompt from me) Ur they need a nice and water to cool them down …. or a swim in the ocean …or a coke. Coke just makes you burp.

6. How do you know that someone has put too much bubble bath in the bath?

You take the plug out

6) Making Inferences from short passages

7. Jane and her friend Sasha went out for a bike ride. They came back later than they were supposed to and Jane was pushing her bike.

Why?

Um – (repeat). Cos um maybe because her pet was in the – you know the way …. Is it a girl or a boy? … her pet was in the way…. My girlfriend has a pet. She has a little puppy and her name is Lexi. (QS) The puppy ---- Kayley. Sometimes I just see her to go play with her… She has an ipad.

10) Strange Stories

White lie

Tom's friend Samia has just had her hair cut. Tom doesn't like it. He doesn't think it suits her. (I don't like it either) He liked her old style better. When Samia asks Tom 'Do you like my new hair style? Tom says 'Yes, it's very nice.'

11) Is it true what Tom said? Yep
12) Why did he say it? Cos it was beautiful. Yep yep yep.

11) Faux pas

Disco

It was the school disco. Jill was in one of the toilets. Karen and Sara came in to the toilets afterwards. Karen said: 'Did you see how Jill was dancing. Wasn't she terrible!' Then Jill came out of the toilet.

1) How does Karen feel? Oh you mean like Karen from sponge Bob? Um happy
2) How does Jill feel? Um – kinda like dancing like toy Barney. (Rpt) Happy
3) Did Karen know that Jill was in one of the toilets? Yeh.

Vase

Mrs Patel has Mrs Brown, an old friend, over for tea. She hasn't seen Mrs Brown for years. Mrs Brown accidentally knocks over a vase and breaks it. Mrs Patel says 'It doesn't matter. It was a wedding present that I never liked anyway. I can't remember who bought it for us." Mrs Brown says 'I bought it for you.'

10) How does Mrs Brown feel? Happy and I think she bought her a home toy from McDonalds. 'My hands in the air just like I do not care'.
11) Did Mrs Patel remember that Mrs Brown had bought her the vase? Um – I fink so.
12) How does Mrs Patel feel? Happy

12) Idioms

"Some sayings mean something different from their words. For example 'give me a hand' doesn't mean 'chop a hand off and give it to me'. It means 'help me please.' (She means – I didn't buyed if for you and maybe your daddy or your auntie or uncle buyed it)What do the following sayings mean? It is okay to say 'don't know' if you do not know what it means."

1. Keep your eyes peeled Um – like closing them. Maybe it might be like a secret surprise.

2. I've got butterflies in my tummy Ur. It means she might be sick like this (demos). I be like that all week.

3. My lips are sealed What's seal mean?

4. Let's hit the road. If you try to hit the road then if you step on it all the time than a car or a bus or a taxi might run over you and you have to go home and have a plaster on it.

5. Put a sock in it Um – Put a sock in the washing machine

6. This is a piece of cake For eating

7. She's over the moon Cos she was put a flag on the top of the moon

8. He's having a whale of a time Whale of the time? Like swimming with it.

9. She got out of the wrong side of bed today Cos it was morning

10. He's driving me round the bend Cos he might be trying to take theirselves to the cinema to see the new Sponge Bob movie.

Appendix xix

Video/DVD Consent Form

Name of Child: ________________________

Name of Parent: ______________________

Parental Consent

I agree to the video of the assessment of my child being used for the following purposes:

1. For the general teaching of SLT students.

Please circle as appropriate Yes/No

2. For wider training purposes (e.g. as part of a conference presentation).

Please circle as appropriate Yes/No

Your confidentiality will be protected at all times. All data will be kept indefinitely. However, you may ask at any time for all or part of the data held about you or your child to be destroyed.

I understand that my confidentiality will be protected at all times and that I may withdraw my consent at any time.

Signed: ______________________________________ (Parent)

Date:___________________

Appendix xx – NW Network Conference presentation

Developing a British Assessment of Inference and Idiom Comprehension

- Anne Hewitt
- Department of Health Professions
- MMU
- a.hewitt@mmu.ac.uk

Background

- PhD study
- developing an assessment of inference and idiom comprehension for 5-10 year old British children.
- 25 years clinical experience as a paediatric Speech and Language Therapist (SLT) shown existing assessments don't cover these areas in enough detail and are not standardised on a British population.

Definitions

- ***Inferential comprehension*** is the ability to link language to a relevant context and to 'read between the lines' when information is not explicitly stated. This ability also comes under the umbrella of 'verbal reasoning'.
- ***Idioms*** involve non-literal understanding of verbal language.
- ***Both essential for successful communication.***

Discrepant performance

- Some children with verbal comprehension difficulties perform at an age appropriate level on assessments of grammatical structures
- ***But*** still have difficulties with inference and idiom comprehension.

Existing assessments

- The main informal assessment of verbal reasoning ability currently used in the UK is the Test for Abstract Language Comprehension (TALC 1 for younger and 2 for older children).
- The TALC (Elks and McLachlan, 2004) is based on the findings of Blank, Rose and Berlin (1978). *The Language of Learning: the preschool years.*
- They studied the language of American 'pre-school' children (up to 6 years in the US) and recorded the different levels of comprehension needed in the classroom setting.

Higher Level Comprehension

- The Blank Model (1978) is used a lot in Australia and is being increasingly used in the UK (as it is covered in the Elklan training).
- It is an old reference but is a seminal work and is a useful framework for comprehension beyond key word levels.
- It covers the ages 3-6 years.
- It has 4 levels of comprehension, progressing from concrete to abstract (single word understanding to verbal reasoning).
- Summary of stages provided in next 6 slides.

Blank Levels

Level 1: Matching Perception:

- Scan for a matching object (**Find one like this**)
- Identify an object by sound/touch (**Show me what you heard/ touched**)
- Name an object heard/ touched/ seen (**What did you hear/ touch/ see?**)
- Imitate a simple sentence (**Say this ..**)
- Remember pictured objects and incidental information (**What did you see?**)

Blank – Level 2

Selective Analysis of Perception:

- Find an object by function (**Find the one we can ride on**)
- Describe a scene (**What's happening?**)
- Recall items and information from a statement (**What things? Who/ what/ where?**)
- Complete a sentence (**Finish this …**)
- Concepts (**Find the one that is X and Y; How are these different?; Name something that is a fruit/ animal**)

Blank Level 3

Reordering Perception:

Now have to evaluate material and ideas:

- Scan for an object integrating verbal and visual information (**Hold up an apple – Find something we could cut this with**).
- Describe What will happen next? From a scene.
- Assume someone else's role (**What might the boy say?**)
- Follow a set of directions (**Put your shoes on, get your coat and meet me at the car**)
- Arrange (4) pictures in the correct sequence.
- Describe a sequence (**How do you make a jam sandwich?**)

Blank Level 3 (cont)

- Generalize about a set of events (**What happened to all of these?**)
- Tell the story from a sequence of pictures.
- Concepts - Identify similarities (**How are these the same/alike?**)
- Concepts – exclusion (**Find the one(s) that we don't wear; What can you write with that is not a pencil?**)
- Concepts – defining words (**What is an X?**)

Blank Level 4

Reasoning about perception

Involves complex problems that require reasoning about what could happen.

- Predicting (**What will happen if…?**)
- Justifying decisions (**If I put the ice cream in the oven would it stay frozen? Why not?**)
- Identifying the cause of an event (**What made it happen? How did the light go off?**)
- Formulating solutions to problems (**This soup is too hot to eat. What could you/he do to cool it down?**)

Level 4 (cont)

- Selecting and explaining means to a goal (**What could we use to pour these beads into this bottle? (a funnel)**.
- Explain the construction of objects (**Why are keys made out of metal and not cloth?**)
- Explain an inference from an observation (**How can you tell that this girl is not happy?**)
- Explain compound words (**Why is this called a button hole?**)
- Explain the obstacles to an action (**Why can't we fit this piece in this puzzle?**)

Blank norms

These norms are based on a very small sample (20 children per year group).

- 60% of 3 year-olds understand level 1 and 2 questions.
- 65% of 5 year-olds understand level 3 and 4 questions.
- Level 4 skills develop between 4½ and 6 years of age.
- Your views on the TALC assessments?
- My aim is to develop an assessment covering verbal reasoning skills beyond 5 years of age up to 10 years, called the HICIT (Hewitt Inferential Comprehension and Idioms Test).

Other assessments

- These include the Liverpool Language Cards; sub-tests of the ACE and the CELF4.
- The Liverpool Language Cards verbal reasoning screen was developed in the 1990s by the Liverpool language unit SLTs. A wide range of Blank level 3 and 4 questions were made into an informal assessment which was trialled on 75 typically developing boys and 71 girls aged 4-7.
- The Assessment of Comprehension and Expression 6-11 (Adams et al, 2001) includes a sub-test on inferential comprehension (the burglar story) and an extension sub-test on non-literal comprehension.

Other assessments

- Clinical Evaluation of Language Fundamentals (CELF-4-UK) (Semel et al, 2006). This is a US assessment but it has also been standardised on a UK population. It covers ages 5 to 16:11 years. The sub-section Understanding Spoken Paragraphs includes some Inferential questions.
- Test of Language Competence – Expanded Edition (Wiig et al, 1989). This US assessment covers ages 5-18:11. It includes sub-sections on making inferences and figurative language.

Pilot to final assessment

- Pilot results: Trialled on 4 ½ to 10 ½ year-olds (10 per year-group). Found -assessment most appropriate for 5 to 9:11 year olds.
- 14 sub-sections of the pilot test reduced to 12 sub-sections
- Final test carried out with 250 children aged 5 to 9:11 from September 2010 to July 2014.
- The assessment is of verbal comprehension but it requires a verbal response. However, the child only needs to have basic expressive language skills to give a correct response.

Pilot to final assessment

- Sections 7 and 8 have accompanying emotions pictures and section 9 has supporting written words.
- Some questions require single word answers and some require a short phrase or sentence.
- Auditory memory demands of the test are mitigated by allowing as many repetitions of the question as requested by the child.
- Harder vocabulary (e.g. enamel) can be defined.
- There are detailed marking criteria for each question, expanded over the course of the study. Some questions have 1 correct answer and some allow for a variety of correct responses.

HICIT Test Construction

The inferential comprehension test questions based on:

- Blank level 3 and 4 questions.
- Examples from existing assessments (using different scenarios).
- Some of the findings on Theory of Mind (emotions questions, mental state verbs, strange stories, faux pas).

The idioms are drawn from resources such as 'Don't take it so literally' and other idiom examples. These have been narrowed down to the 20 easiest to understand idioms.

12 Sub-tests (140 questions)

1. Simple deductions (8 Qs)
2. Why questions (12 Qs)
3. Making predictions (6 Qs)
4. Formulating solutions (12 Qs)
5. Explaining inferences (12 Qs)
6. Making inferences from short passage (10)
7. Situation based emotions (16 Qs)
8. Belief based emotions (8 Qs)
9. Mental state verbs (12 Qs)
10. Strange stories (12 Qs)
11. Faux pas (12 Qs)
12. Idioms (20 Qs)

- Sections 9-12 too difficult for the 5 year-olds so they were only tested on sections 1-8 (84 questions).

Examples of Questions and Answers

1) Deductions:

4. I am clothes. I am often made from wool. You put me on your hands. I keep your hands warm when it is cold.

What am I? (gloves/ mittens/ muff) A squirrel (boy 5:01)

6. You find me inside and outside buildings. I have a frame. You can see through me. If you throw a hard ball at me I could smash.

What am I? (window) A bus stop (girl 6:02)

Question examples

2) Why Questions:

8. Why is a pan made out of metal not chocolate?

- Because if you ate it you won't be able to draw on it (boy 5:07) (thought 'pen'?)

12. Why do we need pedestrian crossings on roads? (NB If the child asks what a pedestrian crossing is you can say: 'Like a zebra crossing')

- The cars stop and let the zebra past if it has escaped from the zoo ! (girl 6:09)

Examples of Questions

3) Making Predictions

1. What would happen if you left a block of ice in the sun? The world will be fire (boy 5:00)

6. What would happen if you had a glass of water full up to the top and then you added 6 ice cubes? It'll displode (girl 5:04)

4) Formulating solutions

4. What could you do if you wanted to eat your soup but it was too hot? Go on a trampoline for 5 minutes (boy 8:07)

12. What could you do if you locked yourself out from your house and nobody else was at home? Wait til the people that own you come back (boy 7:02) (marked correct)

Examples of Questions

5) Explaining Inferences (How do you know?)

N.B Prompt all 'Because they tell you' answers in this section.

3. How do you know that someone is angry? Their eyebrows go really sideways (girl 8:05) (correct)

9. How do you know what's inside a box of something at the supermarket? Use your senses. You could see a tail if it's a cat (girl 7:08)

6) Making Inferences from short passages

5. Dan's dad took him to somewhere outside. He pushed Dan up and down on something and pushed him round and round on something else. Dan climbed up some steps and went down something.

Where were Dan and his dad? In the basement (boy 8:00)

Examples of Questions

10. Dad bought a chicken to cook for Sunday dinner but the family ended up eating just vegetables and gravy.

What happened?

Maybe the chicken got burnt. Maybe it was an actual live chicken and it ran away. Or maybe it was a BBQ outside and a fox took it away when dad was playing swing-ball with the kids and mum was inside making vegetables and gravy (girl 7:08) (correct but long-winded!)

Examples of Questions

7) Situation-based emotions inferences (Based on Howlin, Baron-Cohen and Hadwin, 1998)

Use the 4 emotions symbols for this section. Before you start check that the child knows what picture depicts what emotion by asking him/her to point to each one as you say them.

Instructions to the child: "For the next section, choose the feeling that you think fits best from: happy; sad; angry; or scared. There might be two possible answers for some of these but choose the one feeling that fits the best."

1. Rachel's pet cat died. How does she feel? (sad)
3. A big fierce dog chased Raj down the road. How does he feel? (scared)
4. Lucy's dad bought her some sweets. How does she feel? (happy)
9. A boy hit Gemma's sister at school. How does Gemma feel?

Happy if she hates her sister or angry (boy 9:06)

Examples of Questions

8) Belief-based emotions inferences (based on Howlin et al 1998)

Just have the 2 symbols for happy and sad visible for this section (i.e fold the emotions sheet in half).

Instructions to the child: "For the next section, choose the feeling that you think fits best from: happy or sad"

Raj wants a toy plane for his birthday but he is getting a teddy.

Raj thinks he is getting a toy plane for his birthday.

- 3) How does he feel? (happy).
- 4) He gets a teddy for his birthday. How does he feel? (sad)

Sad – but he might be a little bit happy cos at least he got something (boy 7:09) (correct)

Examples of Questions

9) Mental state verb inferences (based on Spanoudis et al 2007)

Use the page with the words Yes, Maybe and No on it. Point to the appropriate word as you give the instructions below.

'I am going to tell you about 12 different people and a door. If you think the door is definitely locked say 'yes'. If you think the door might or might not be locked say 'maybe'. If you think the door is definitely NOT locked say 'no'.

1. Karen knows that the door is locked. Is the door locked? (yes)
2. Jack thinks that the door is locked. Is the door locked? (maybe)
4. Shazia forgets to lock the door. Is the door locked? (no)

Examples of Questions

10) Strange Stories (Based on Happe, 1994)

NB. If the child answers 'yes' to the 'Is it true?' question probe 'why?

Pretend

John and Rashid are playing spacemen. John picks up a bowl and puts it on his head. He says: 'This bowl is a space helmet.'

1) Is it true what John says?

2) Why does he say it?

Because he's imaginating (girl 7:02)

Examples of Questions

11) Faux pas (based on Baron-Cohen et al, 1999)

N.B It is OK to remind a child who specific people are if they ask.

Clothes shop

Laura worked as a clothes shop assistant. A lady with a fat tummy came in and tried on a dress. Laura said 'That looks lovely and you will still be able to wear it when you have had the baby.' The lady said 'I'm not pregnant.'

- 7) Did Laura know that the lady was not pregnant?
- 8) How does Laura feel?

Sad and a bit hysterical (boy 8:03) (correct)

- 9) How does the lady feel?

Very cross and that she needs to go to the gym (girl 8:05)

Examples of Questions

12) Idioms

Instructions to the tester: Say all of these idioms with a fairly neutral tone of voice and facial expression and do not gesture, so that you do not give away the meaning.

Introduce the section by giving the following example:

"Some sayings mean something different from their words. For example 'give me a hand' doesn't mean 'chop a hand off and give it to me'. It means 'help me please.' What do the following sayings mean? It is okay to say 'don't know' if you do not know what it means."

1. Keep your eyes peeled.

It means that there's rats or monsters around (boy 9:05)

Idioms

2. I've got butterflies in my tummy.
You have a dizzy tummy (boy 7:04)
12. It's not my cup of tea
It's not your teabag (boy 7:05)
I don't really like it – I prefer black coffee (girl 9:06)
16. I'm tied up at the moment.
Many children said 'tired'.
Someone ties them up in prison after getting in a matrix van with other people (girl 6:05)
19. It goes in one ear and out the other.
A piece of string (boy 6:06)
20. You've hit the nail on the head.
You've give me a headache (boy 7:09)

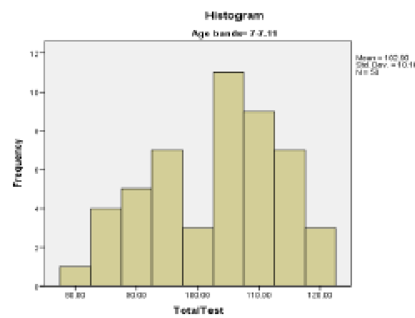
Assessment results

- **Final Total Scores: 5-5:11 year-olds**
- Only sections 1-8 completed with this age group (sections 9-12 were too difficult for this age group in the pilot study)
- Total maximum possible score is **84.**
- **Boys aged 5-5:11: Mean score - 48.12 Range: 32-64**
- **Girls aged 5-5:11: Mean score - 48.48 Range: 27-67**
- **Combined (boys and girls): Mean – 48.3 Range: 27-67**

Total Test Scores: 6:00 - 9:11

Age Band	Sex	Mean (/140)	Range
6:00-6:11	Boys	89.48	72-123
6:00-6:11	Girls	86.72	51-108
7:00-7:11	Boys	99.24	81-120
7:00-7:11	Girls	106.36	80-118
8.00-8:11	Boys	105.72	90-122
8:00-8:11	Girls	108.28	85-126
9:00-9:11	Boys	118.8	100-132
9:00-9:11	Girls	120.44	100-134

Normal distribution



Sub-section 1) Deductions scores

Age	Mean	5%Trimmed Mean	Median	Variance	Standard Deviation	Min/Max (Range)	Interq Range	Skewness	Kurtosis	K-S* Sig.
5-5:11	4.48	4.5	4.5	2.867	1.69320	1-7 (6)	3	-.122	-1.148	.001
6-6:11	5.48	5.4778	5.5	2.214	1.48791	3-8 (5)	2.25	0.79	-.803	.021
7-7:11	6.44	6.4778	7	1.313	1.14571	3-8 (5)	1.25	-.525	.058	.000
8-8:11	6.72	6.7889	7	1.308	1.14357	4-8 (4)	2	-.612	-.424	.000
9-9:11	7.061	7.1689	7	1.142	1.06865	3-8 (5)	2	-1.407	2.906	.000

Use of HICIT with children with SLCN

- A 5½ year-old child with a high functioning Autistic Spectrum Disorder (ASD) was assessed as part of the test standardisation process (not included in the collation of the scores).
- His language profile:
- Good comprehension of grammatical structures. On the TROG(2) aged 5½ he obtained an Age Equivalent of 6:06 (Standard Score 113; 81st percentile).
- Scored age appropriately on the information and grammar sections of the Renfrew Action Picture Test.

Sub-test scores for child with ASD

1. Simple deductions (5/8)
2. Why questions (0/12)
3. Making predictions (0/6)
4. Formulating solutions (3/12)
5. Explaining inferences (2/12)
6. Making inferences from short passage (4/10)
7. Situation based emotions (9/16)
8. Belief based emotions (3/8)

Quantitative and Qualitative responses

- The child with ASD scored 26/84 – well below the mean of 48.12 for 5-5:11 year-olds and well below the lowest score in the range for boys (32-64).
- This is significant in view of his more than age appropriate comprehension of grammatical structures (on the TROG 2).
- It indicates a specific problem with his inferential comprehension.
- His qualitative responses are also very interesting. He gave some very unusual responses to some of the test questions! Spot his special interests.

DVD Examples from other children

Ethan (PLI): Aged 10:02 – TROG A.E 6 (-9) years

- HICIT Total Test Score: 50 (mean for 6-6:11 is 88)
- He scored well below the mean in ALL sections apart from deductions; formulating solutions and situation-based emotions.
- DVD clips illustrating his difficulties. Spot his special interest.

Leonardo (ASD): Aged 9:10 – TROG2. Passed 19 blocks. A.E over 12 years.

- HICIT Total Test Score: 104 (mean for 9-9:11 is 120. He scored at the lowest end of the range)
- His main difficulties were with belief-based emotions, faux pas and idioms.
- DVD clip illustrating his qualitatively different responses.

Differential diagnosis

- The above examples shows that the HICIT could be useful as part of a battery of assessments for diagnosing the receptive language difficulties characteristic of high functioning ASD and Pragmatic Language Impairment.
- Once completed, the assessment could also be trialled with people with other kinds of receptive language difficulties (for example Specific Language Impairment or as a result of epilepsy or brain injuries) to see if it can help differential diagnosis.
- Your thoughts on if it could be a useful clinical tool?

References

Adams, C., Cooke, R., Hesketh, A. & Reeves, D. (2001) *ACE – Assessment of Comprehension and Expression 6-11*. NFER NELSON. Windsor.
Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R. and Plaistea, K. (1999): *Recognition of faux pas by normally developing children and children with Asperger Syndrome or high-functioning autism*. Journal of Autism and Developmental Dirorders 29(5):407-18.
Barrett, M., Huisingh, R., Zachman, L., Blagden, C. and Orman, J. (1992) *The Listening Test*. Linguisystems Inc.
Bishop, D.V.M (1997): Uncommon Understanding: Development and disorders of language comprehension in children. Psychology press. Hove.
Bishop, D.V.M (2003) *Test for Reception of Grammar 2*. Pearson Assessment. PsychCorp.
Blank, M., Rose, S. & Berlin, L. (1978) *The Language of Learning: the preschool years*. Grune & Stratton. New York.
CELF-4, Semel, E. (2003)
Elklan Resources: www.elklan.co.uk
Elks, L. and McLachlan, H. (2004) *The Test of Abstract Language Comprehension (1 and 2)*. Elklan Publications, Cornwall.

References (cont)

Happe, F.G.E. (1994). *An advanced test of theory of mind: Understanding of story character's thoughts and feelings by able autistic, mentally handicapped and normal children and adults*. Journal of Autism and Developmental Disorders, 24, 129-154.
Howlin, P., Baron-Cohen, S. and Hadwin, J. (1998): *Teaching children with Autism to Mind-Read: A Practical Guide for Teachers and Parents*. Wiley Publishing.
Semel, E., Wiig, E.H., and Secord, W. (2006). *Clinical Evaluation of Language Fundamentals. 4th Edition. UK.* Pearson Education Ltd.
Spanoudis, G., Natsopoulos, D and Panayiotou, G. (2007). *Mental verbs and pragmatic language difficulties*. International Journal of Language and Communication Disorders. Vol 42, 4, p.487-504
Wiig, E.H., Secord, W. (1989) *Test of Language Competence*. The Psychological Corporation. Harcourt Brace and co.
a.hewitt@mmu.ac.uk