Please cite the Published Version

Eye gaze and L2 speakers' responses to recasts: A systematic replication study of

McDonough, Crowther, Kielstra, and Trofimovich (2015)

Kim McDonough, Pavel Trofimovich, Phung Dao, & Dato Abashidze

Concordia University

Address for correspondence

1455 de Maisonnueve Blvd W

Education Department, FG 6.151

Montreal, QC H3G 1M8 Canada

Email: kim.mcdonough@concordia.ca

Phone: 514-848-2424 ext 5168

Author biographical notes

Kim McDonough is a Professor of Applied Linguistics in the Education Department at Concordia University. Her current research interests include joint attention during task-based interaction and collaboration in L2 writing.

Pavel Trofimovich is a Professor of Applied Linguistics in the Education Department at Concordia University. His research focuses on cognitive aspects of second language processing, second language speech learning, sociolinguistic aspects of second language acquisition, and the teaching of second language pronunciation.

1

Phung Dao recently completed his PhD in Education with a specialization in Applied Linguistics at Concordia University. His research interests include learner engagement in peer interaction, and task-based language teaching.

Dato Abashidze is a Horizon Postdoctoral Fellow in Applied Linguistics at Concordia University. His research focuses on the contributions of linguistic, visual and cognitive factors in the comprehension of spoken language and learning.

2

Abstract

To confirm the role of social factors in mediating cognitive processes, this systematic replication study seeks to extend the generalizability of an exploratory study (McDonough, Crowther, Kielstra, & Trofimovich 2015) that reported a positive association between eye gaze and second language (L2) speakers' responses to recasts. For this replication, L2 English speakers ($N = 74$) carried out communicative tasks with research assistants who provided recasts in response to non-targetlike forms while both interlocutors' eye gaze behavior was tracked. Transcripts were analyzed for the occurrence of recasts in response to different error types, recast length, and L2 speaker responses. Eye gaze length for the research assistants when producing the recast move and the L2 speaker when responding to the recast were obtained in seconds, and mutual gaze (i.e., simultaneous looking) was included as a binary eye gaze variable. A logistic regression model confirmed the findings of McDonough et al. (2015), with both L2 speaker and mutual eye gaze predictive of targetlike responses; however, the effect of L2 speaker's eye gaze duration was in the opposite direction as compared to the initial study. Implications are discussed in terms of understanding the role of eye gaze in face to face interaction.

Key words: Eye gaze, recasts, responses to recasts

Eye gaze and L2 speakers' responses to recasts: A systematic replication study of

McDonough, Crowther, Kielstra, and Trofimovich (2015)

1. Introduction

Responding to repeated calls for greater attention to the social turn in second language (L2) acquisition research (Firth & Wagner 1997; Block 2003; Batstone 2010; Hulstijn et al. 2014), interaction research has expanded beyond a largely cognitive orientation to incorporate the importance of social factors in mediating cognitive processes. Researchers have argued that interpersonal and contextual factors, perceived as micro-social factors (Saville-Troike & Barto 2017), can either facilitate or inhibit L2 learning through interaction. Various social factors, such as group composition (e.g., pair or small group) and interpersonal relationships between learners (e.g., familiarity, attitude, and shared linguistic or cultural background) have been shown to influence the degree to which L2 speakers engage in interactional processes, such as producing accurate language, providing and receiving feedback, attending to each other's corrections, and reformulating their non-targetlike forms (e.g., Philp & Duchesne 2008; Philp & Mackey 2010; Dobao 2014; Choi & Iwashita 2016; Poteau 2017). Additionally, social relations reflected through pair/group dynamics (Storch 2002; Storch & Aldosari 2012; Sato & Viveros 2016) and learners' positioning in interaction (Young & Tedick 2016) affect the quantity and quality of collaborative discussions about language form.

Among the various interactional features believed to promote learning, feedback has received considerable attention from L2 researchers (see Plonksy & Brown 2014, for a review). Numerous meta-analytic studies have examined the amount, type, characteristics, and

4

effectiveness of interactional feedback that occurs in both classroom and laboratory settings (e.g., Russell & Spada 2006; Mackey & Goo 2007; Li 2010; Lyster & Saito 2010; Plonsky & Gass 2011; Miller & Pan 2012), with overall findings indicating benefits for interactional feedback, but with variation across settings, proficiency levels, linguistic targets, and specific feedback types. The effectiveness of recasts in particular has been widely debated (Goo & Mackey 2013; Lyster & Ranta 2013), with questions raised concerning the limited amount of repair they elicit and learners' difficulty in perceiving them as corrective. Reflecting these debates, numerous research studies have investigated which recast features impact the extent to which L2 speakers reformulate their initially non-targetlike utterances, which includes linguistic target, length, stress, and intonation (Lyster 1998; Mackey, Gass, & McDonough 2000; Philp 2003; Carpenter, Jeon, MacGregor, & Mackey 2006; Loewen & Philp 2006; Sheen 2006; Egi 2007).

In addition to such recast features as length and linguistic target, another factor that might play a role in whether L2 speakers reformulate their non-targetlike forms is the presence of nonverbal communicative cues, such as gestures, facial expressions, and eye gaze. The possibility that nonverbal visual cues may help L2 speakers recognize feedback has been acknowledged by researchers working in both classroom and laboratory settings (Lyster, 1998; Gullberg 2006; Long, 2007; Gullberg 2010). Classroom-based studies have largely focused on instructors' general use of gestures in L2 classrooms (e.g., Allen 2000; Lazaraton 2004; Smotrova & Lantolf 2013), but have also examined their use of visual cues during focus on form and feedback episodes specifically (Davies 2006; Faraco & Kida 2008; Wang & Loewen 2016).

5

Over half of the corrective feedback in Wang and Loewen's study of ESL teachers was

accompanied by visual cues, most frequently nodding, head shakes, and pointing. For recasts and

clarification requests, Davies found that ESL speakers' repair rates were higher when the

feedback moves were accompanied by visual cues, while Faraco and Kida highlighted the

importance of gaze cues in negotiation of meaning sequences. Nevertheless, experimental

research has not found that recasts with gestures are more likely to be noticed than recasts

without visual cues (Nakatsukasa 2016), or that raters report using visual cues to identify recasts

(Carpenter et al. 2006).

Focused exclusively on visual cues in the form of eye gaze, McDonough and colleagues

(McDonough, Crowther, Kielstra, & Trofimovich 2015) carried out an exploratory study to

examine whether eye gaze predicted L2 speakers' targetlike responses to recasts. Motivated by

the joint attention research (Moore & Dunham 1995), they investigated whether mutual gaze

(i.e., when both interlocutors look at each other) facilitated L2 speakers' targetlike responses.

Full-time ESL students ($N = 20$) carried out four communicative tasks with research assistants

who provided feedback in response to their non-targetlike forms. Both interlocutors' eye gaze

was captured using the faceLab 5 eye-tracking system, and their interactions were audio

recorded. Transcripts were analyzed for the occurrence of all types of interactional feedback, but

following data pruning, only recasts provided in response to grammatical errors were included in

the regression analysis. The outcome variable was targetlike responses, and the predictor

variables were recast features (length, intonation, stress) and eye gaze measures (mutual gaze,

RA eye gaze while delivering the recast, L2 speaker eye gaze when responding). The regression

6

model was significant, and two variables—mutual gaze and L2 speaker eye gaze—were

significant predictors of targetlike responses.

Despite the positive findings for eye gaze, the authors cautioned that replication studies

were needed due to the small sample size, homogeneity in the sample (80% were Mandarin

speakers of English), and technical issues with the positioning of the video cameras (behind the

interlocutors). In response to their call for replication, this systematic replication study seeks to

test the generalizability of the finding that eye gaze is associated with L2 speakers' targetlike

responses to recasts. Because this study was conceived as a systematic replication, the methods

of the initial study (McDonough et al. 2015) are duplicated, but some key variables have been

altered, specifically the participants (larger sample, more diverse language backgrounds, higher

proficiency), the experimental setup (improved placement of webcams), and the interactive tasks

(only two tasks, with one more academically oriented). The research question was: *Does eye*

gaze predict L2 speakers' responses to recasts? Based on the findings of the initial study, it was

predicted that mutual eye gaze and the duration of L2 speaker eye gaze would predict targetlike

responses to recasts.

2. Method

2.1 Participants

The participants were 74 English L2 speakers (43 women, 31 men) enrolled in undergraduate (n

$= 48$) and graduate ($n = 26$) degree programs at an English-medium university in Montreal,

Canada with a mean age of 25.6 years ($SD = 4.9$). A power analysis carried out using the pwr

function in R (R Core Team, 2016) revealed that a minimum sample of 48 participants was

needed to detect a medium effect size using four predictor variables in a regression model, with alpha set at .05 and power estimated at .80. This analysis confirmed that the initial study was underpowered and that the current sample of 74 participants was sufficient for the current replication study. The participants had studied English for a mean of 10.8 years (SD = 5.4) and resided in Canada for 3.3 years (SD = 3.2). They came from a variety of first language (L1) backgrounds including Mandarin (24), Farsi (14), Vietnamese (13), Arabic (8), French (6), Spanish (4), Russian (2), Hindi, Swedish, and Ukrainian (1 each). As for their English proficiency, the participants reported IELTS scores ranging from 6.0 to 7.5 (M = 6.5, SD = .4.) or TOEFL iBT scores ranging from 75 to 100 (M = 86.6, SD = 8.6). These scores place the students in the B2–C1 range of the Common European Framework of Reference. Compared to the sample in the initial study (N = 20), the current participants were approximately four years older, had resided in Canada about two years longer, and had higher English proficiency scores (B2–C1 versus B1). In addition, whereas the sample in the initial study was predominantly Mandarin L1 speakers (80%), Mandarin speakers accounted for only 32% of the current sample. The current sample also included speakers from 10 different language backgrounds, compared to only three in the initial study. By recruiting a larger and less homogenous participant sample it was thus possible to minimize potential language- and culture-specific effects on study outcomes.

2.2 Materials

The materials consisted of interactive activities in the form of an interview task and a discussion task based on TED talks. The interview task, which was used in the initial study, was based on 12 cards that listed the following topics: travel and adventure; love, dating, and marriage; dreams

and goals; social media; science and technology; music and entertainment; jobs and occupations;

the environment; inspirational people; moving to a different country; restaurants and fine cuisine;

and stress and relaxation. The TED talk activity was created for use in the replication study due

to the higher proficiency and greater academic orientation of these participants as compared to

the previous sample. It consisted of a discussion about one of two TED talks about green

building or web tracking (approximately 6.5 minutes long) selected by the participant. The

activity consisted of warmup questions, watching the TED talk, summary of the presenter's main

points, and an exchange of opinions about the topic and themes raised by the presenter.

2.3 Procedure

As in the initial study, the participants carried out the research activities during individually

scheduled sessions (120 minutes) with research assistants serving as either the interlocutor or the

eye-tracking assistant. Across the sessions, two research assistants (henceforth, interlocutors)

were randomly assigned to interact with an equal number of participants. These two interlocutors

(one male, one female) were French-English bilinguals (native speakers of French and highly

proficient in English) in their late 20s enrolled in an MA program in applied linguistics with L2

teaching experience.

The participants first completed a consent form (5 min) and then underwent a brief four-

point calibration for the eye-tracking equipment (15 min) with the eye-tracking assistant. The

faceLAB 5 eye-tracking system was used to capture the eye movements of the RA interlocutor

and participant, who were seated at a table opposite each other with four cameras positioned on

two stereo heads in the middle of the table, so that two cameras tracked and recorded the eye

gaze of each person. Two Logitech webcams were placed next to the four cameras to record the scene, which included the head and torso of each person. The placement of the webcams in the middle of the table ensured uninterrupted, clear recording of the visual scene. In the initial study, these webcams had been located behind the interlocutors, which caused some technical problems (and therefore loss of information) due to inadvertent contact and blocking of the scene. Together these cameras integrated the eye movement and field-of-vision data, specifically where in the scene (as depicted visually by a green dot in the field of vision) each person looked. The cameras were connected to two synchronized DELL Latitude E5520 laptops recording both interlocutors' eye gaze behavior.

After calibration, the eye-tracking assistant left the room and monitored the eye-tracking equipment in an adjacent room using a screen share application. Then the participant and the interlocutor carried out the interview task (15 min) and the TED talk discussion (25 min), during which the interlocutor provided recasts when the participant produced non-targetlike forms. As in the initial study, the types of errors to be recast or the number of recasts to provide were not decided a priori. Instead, the interlocutors were instructed to provide recasts whenever they felt it was appropriate to do so. In other words, as in the initial study, the interlocutors were asked to interact and give feedback as naturally as possible and to adjust their feedback if a participant was perceived to be worried or annoyed, with the overall goal for the interlocutor to maintain a primary focus on communication. Thus, the absence of interlocutor training on recasts or explicit instruction about the type and frequency of recasting was consistent with the focus of the initial study and its replication on the role of visual cues in naturalistic meaning-focused interaction.

The interaction between a participant and the interlocutor while carrying out the tasks

was recorded using a digital audio-recorder. After completing the interactive activities, the

participant went to an adjoining room where the eye-tracking assistant debriefed each participant

about their perceptions of the tasks, equipment, and the interlocutor (10 min). As part of the

same data collection session, the RA interlocutor narrated a close call event at the beginning of

the session (20 minutes) that the participants later recalled (5 min), and the participants

completed questionnaires about their motivation and acculturation (25 min); however, their

performance on these activities is not included in the current study and is reported elsewhere

(Authors, in review). All materials for the initial and the current replication study are available

via IRIS (http://www.iris-database.org) and Open Science Framework (https://osf.io).

2.4 Coding and analysis

The audio recordings were transcribed and verified by research assistants. As in the initial study,

the interaction data were coded for recasts, which were operationalized as more targetlike

reformulations of participants' erroneous forms. The type of error targeted in each recast was

classified as being about grammar, lexis, phonology, or multiple errors involving different forms.

Grammar errors included sentence formation, word order, morphological, and syntactic errors;

lexical errors involved incorrectly used words, including semantically inappropriate words, false

cognates, or otherwise imprecise vocabulary choice; phonology errors included phonemic

substitutions, deletions, and insertions as well as word stress errors. Sample errors for each coded

category, along with sample recasts targeting these errors, are provided in Table 1.

11

Table 1 *Sample Error Types Followed by Recasts*

Error type	L2 speaker' utterance	Recast
Grammatical	in Europe the strongest country is German again	yeah yeah Germany
Lexical	And everybody was so uh intensive	Everybody was so stressed?
Phonological	Chimical factory	Chemical factory?
Multiple	I prefer to swimming in swimming pool	To swim in a swimming pool

The length of the recasts was coded following Sheen (2006) as short phrase (one content word), long phrase (two or more content words), or clausal (subject and tensed verb). Because the initial study found no relationships between responses to recasts and intonation or stress, these recast characteristics were not coded. This methodological decision also allowed for including a larger number of recast episodes in analysis, thus increasing its power, than otherwise would be possible if finer grained categories of recast characteristics had been used. Finally, as the binary outcome variable, participants' responses to recasts were coded as *targetlike response* if they produced the form recast by the RA or *no reformulation* if they simply acknowledged the recast (e.g., *yeah*, *uh huh*, *right*), repeated the original error, or continued the conversation. Recasts where participants did not have an opportunity to respond because the interlocutor continued speaking were not included in the dataset. An independent rater coded a

subset of the data (20%), with interrater reliability obtained through Pearson correlations for the

number of recasts (.95) and Cohen's kappa for error type (.82), length (.81), and L2 speaker

responses (.87).

The eye gaze data were coded manually in 40 millisecond frames for the occurrence of

mutual gaze (i.e., present/absent), which was defined as the occurrence of simultaneous looks to

both interlocutors' faces during the recast. In other words, mutual gaze occurred when the RA

interlocutor was looking at the participant's face at the same time that the participant was

looking at the RA's face. Occurrence of a look was determined by using the analysis program

Captiv which allowed for simultaneous viewing of the recorded scene videos, which featured a

moving green dot over the field of vision to indicate eye movement. The interlocutors' fixations

were manually coded frame-by-frame to determine in which area of the scene they had fixated.

In addition to the binary coding of mutual gaze, the RA interlocutor's eye gaze duration to the L2

speaker while delivering the recasts was coded in seconds, as was the L2 speaker's eye gaze to

the RA when responding to the recast. Although eye gaze durations are typically coded in

milliseconds (see Conklin & Pellicer-Sánchez, 2016), the use of seconds as a unit of analysis was

convenient because it allowed for a reader-friendly interpretation of durations. A look began the

moment that the speaker's gaze first landed on the face of their interlocutor, and ended the

moment their gaze left their speaking partner's face. Only the initial pass was considered for

analysis, with no minimum benchmark for length determined a priori. A subset of the data (20%)

was coded with interrater reliability calculated as Pearson correlations for the RA interlocutor

(.89) and L2 speaker eye gaze duration (.91) and as Cohen's kappa for mutual gaze (.90).

3. Results

3.1 Interaction data

A total of 1,418 recasts occurred when the L2 speakers carried out the two interactive tasks with the RA interlocutors, which represents a mean of 13.8 recasts per speaker ($SD = 10.2$). As was the case in the initial study, the L2 speakers produced grammatical errors most frequently (69%), followed by lexical (23%), phonological (7%), and pragmatic or multiple errors (1%). As in the initial study, we excluded phonological, pragmatic, and multiple errors from further analysis due to their low frequency. However, whereas the initial study focused exclusively on grammatical errors, the current dataset included a sufficient number of lexical errors to examine L2 speakers' reactions to recasts as a function of error type (grammatical vs. lexical). In terms of the length of the recasts, clausal recasts were most frequent (46%), followed by short (33%) and long (21%) phrases. Because long phrases included recasts that were as short as short phrases (e.g., *big data?*) and as long as clauses (e.g., *the first one about green houses?*) they were excluded from the regression analysis.

A total of 957 recasts provided in response to grammatical and lexical errors were selected for inclusion in the logistic regression model, of which 588 (61%) were clauses and 369 (31%) were short phrases. In terms of the participants' responses to the recasts, 73% (701) of their responses contained no reformulation and 27% (256) were targetlike responses. By comparison, this larger dataset yielded considerably more recasts of two different lengths in response to two error types and contained more targetlike responses than the initial study, which had only 276 clausal recasts targeting grammatical errors with 51 targetlike responses.

3.2 Predicting responses to recasts

A logistic regression was carried out with responses to recasts as the binary outcome

variable, with 1 coded as targetlike response and 0 coded as no reformulation. The predictor

variables included in the model were error type (binary: grammatical or lexical), recast length

(binary: clause or short phrase), mutual gaze (binary: no or yes), and duration of RA and L2

speaker eye gaze in seconds (both numeric). For the binary variables of error type, recast length,

and mutual gaze, the reference categories were grammatical errors, clauses, and no mutual gaze,

respectively. These data were considered to be appropriate for logistic regression because both

continuous predictor variables demonstrated a linear relationship with the (logit transformed)

outcome variable (Hosmer & Lemeshow, 1989), because the coded data involved independent

observations (i.e., no datapoint fit within more than one coded category), and because the

obtained models showed little evidence of multicollinearity (see below). No data transformations

were performed prior to running the analyses. The frequency counts (for binary variables) and

descriptive statistics (for continuous variables) for each level of the outcome variable are

provided in Table 2 and Table 3, respectively.

Table 2 *Frequency of Responses to Recasts by Binary Predictors*

Predictor category	Targetlike response ($n = 256$)	No reformulation ($n = 701$)
Lexical errors	155	294
Grammatical errors	101	407
Short phrase	69	170
Clause	111	395
Mutual eye gaze: Yes	161	451
Mutual eye gaze: No	19	114

Table 3 *Means and Standard Deviations (in Seconds) for Responses to Recasts by Continuous*

Predictors

Predictor category	Targetlike response ($n = 256$)	No reformulation ($n = 701$)
RA interlocutor eye gaze	1.07 (0.52)	1.29 (2.03)
L2 speaker eye gaze	0.91 (0.57)	1.19 (2.32)

The predictors were entered into the model using hierarchical regression so that variables

identified as influencing the outcome variable in previous research were entered first, followed

by new predictors. Based on the findings of the initial study, mutual eye gaze and L2 speaker eye

gaze duration were entered first, followed by error type and recast length, which have been

shown to influence L2 speakers' responses to recasts (Mackey et al. 2000; Sheen 2006). Finally,

RA eye gaze duration was entered last as eye gaze has been associated with requests for repair

(Rossano, Brown, & Levison 2009), but it was not identified as significant predictor in the initial

16

study. As shown in Table 4, the first model with mutual gaze and L2 speaker eye gaze was

significant, but the second model with error type and recast length led to a significant change and

higher Nagelkerke R^2 value (from .034 to .064). In contrast, the third model with RA eye gaze

led to neither a significant change in the model nor an increase in the Nagelkerke R^2 value.

Therefore, the best fitting model included four predictor variables: mutual gaze, L2 speaker eye

gaze, error type, and recast length.[1] In terms of assumptions and model fit, tests of

multicollinearity showed that the model had no tolerance values below 0.2 (0.29–0.96) and no

VIF values above 10 (1.04–3.47). Analysis of the residuals indicated good model fit (only 4% of

the cases had standardized residuals greater ±2) and an absence of cases with undue influence on

the model (Cook's distance and DfBeta values were less than one).

Table 4 *Summary of Hierarchical Regression Models for Targetlike Response*

Blocks	χ^2	df	p	R_N^2	$\Delta\chi^2$	df	p
1 Mutual gaze & L2 speaker eye gaze	22.40	2	.001	.034			
2 Error type & recast length	43.22	4	.001	.064	20.82	2	.001
3 RA eye gaze	43.30	5	.001	.064	.08	1	.777

Turning to the predictor variables, mutual eye gaze, L2 speaker eye gaze, and error type

were significant predictors of targetlike responses (see Table 5). Mutual eye gaze increased the

odds of a targetlike response by a multiple of 2.13, which is equal to an increase in the

probability of a targetlike response from 9% to 18%, while a change in error type from

grammatical to lexical increased the odds of eliciting targetlike response by a multiple of 2.47 (to

17

20%). However, longer L2 speaker eye gaze decreased the odds of eliciting targetlike response

by 0.61. Recast length (clausal or short phrase) was not a significant predictor.

Table 5 *Summary of Predictor Variables for Regression Model with Blocks 1 and 2*

Predictors	*B*	*SE B*	*Exp(B)*	95% CI	*p*
Mutual eye gaze	.76	.26	2.13	[1.29, 3.54]	.003
L2 speaker eye gaze	−.49	.14	0.61	[0.47, 0.80]	.001
Error type	.90	.26	2.46	[1.48, 4.10]	.001
Recast length	−.29	.26	0.75	[0.45, 1.25]	.270
Constant	−.91	.22	0.40		

The positive relationship between mutual gaze and targetlike response confirms the

findings of the initial study (McDonough et al., 2015), and extends the role of mutual gaze to

recasts of different lengths in response to more varied errors types. However, the negative

relationship between L2 speaker eye gaze and targetlike responses is contrary to the earlier

finding of a positive relationship. In other words, whereas the initial study found that longer

looks to the RA interlocutor while responding to the recast increased the odds of responding in a

targetlike manner, here longer looks decreased those odds.

4. Discussion

Conceptualized as a response to calls for researchers to address social factors in L2 learning, our

initial study (McDonough et al. 2015) and this systematic replication extend prior work on

recasts by considering the role of visual cues in the form of eye gaze. The current results

replicated the main finding of our initial research, namely, that the presence of mutual gaze between the speaker and the interlocutor is associated with targetlike responses to recasts. To the extent that odds ratios are comparable across studies, mutual gaze was associated with a slightly larger odds ratio in the initial study (3.84, or a change from 9% to 28%), compared to the odds ratio obtained here (2.13, a change from 9% to 18%). However, the current results failed to reproduce the obtained positive relationship between the length of the speaker's eye gaze and targetlike responses, instead yielding a negative association (odds ratio of 0.61, compared to the odds ratio of 3.42 in the initial study). The current results extended the generalizability of the initial study (which targeted only clausal recasts in response to grammatical errors) by showing that L2 speakers' responses to recasts is affected by linguistic target. Recasts focusing on lexis eliciting more targetlike responses compared to those focusing on grammar, with an odds ratio of 2.46. However, responses appear independent of recast length as targetlike responses to both short and clausal recasts were similar, with an odds ratio of 0.75. Our research thus broadens the scope of prior work on recasting by demonstrating that similar to individual and contextual variables, such as speaker's anxiety and working memory, interlocutor's proficiency, and task complexity and scope (e.g., Trofimovich, Ammar, & Gatbonton 2007; Sheen, 2008; Revesz 2012; Goo 2016), visual cues through mutual eye gaze also affect responses to recasts.

Seen within the context of the initial study and this replication, the significant role of mutual eye gaze can be understood as both interlocutors engaging in the co-construction of a joint interpretation, which likely involved L2 speakers becoming sensitive to the fact that a conversationally provided recast serves a repair function. In essence, an interlocutor's eye gaze

may place pressure on speakers to reformulate their problematic utterances (e.g., Bavelas,

Coates, & Johnson 2002), but speakers must engage in joint attention to increase the likelihood

of a targetlike response (Goodwin 1981; Richardson & Dale 2005). It is therefore not surprising

that—both here and in the initial study—the interlocutor's gaze alone was not predictive of

targetlike response rates, because the interlocutor's visual cues must be acknowledged by the

speaker, ideally by returning a gaze. In this sense, mutual gaze can be seen as a nonverbal,

socially constructed cue increasing the salience of conversational recasts as a feedback technique

(e.g., Ziegler 2018), one that depends on the willing participation of both interlocutors. Although

multiple culture-specific differences in the use of visual cues have been attested (e.g., Rossano et

al. 2009), there nonetheless appears to be remarkable similarity in how humans use visual cues

across languages and cultures, particularly in cases of non-understanding (Floyd, Manrique,

Rossi, & Torreira 2016). That the positive role of mutual gaze remained significant in this larger

dataset comprised of speakers from multiple linguistic and cultural backgrounds further confirms

that mutual gaze might have value in increasing the salience of recasting.

Whereas mutual gaze was associated positively with targetlike responses, the length of

the L2 speakers' gaze to the RA interlocutor had a negative association, which contrasts with the

positive association reported earlier (McDonough et al. 2015). In essence, unlike in the initial

study, shorter looks to the interlocutor were linked to greater incidence of targetlike responses in

the current dataset. While at first glance these results seem contradictory, when considered

together, they suggest a possible temporal threshold beyond which L2 speakers' eye gaze no

longer positively predicts the extent to which recasts are associated with targetlike responses. In

the initial study, the positive relationship between L2 speakers' gaze duration and targetlike responses was based on gaze durations that were relatively brief ($M = 0.39$ seconds, $SD = 0.46$). In contrast, recasts in this replication study elicited looking behaviors that were on average twice longer ($M = 0.91$ seconds, $SD = 0.57$). It is possible that eye gaze cues of a short duration, such as a brief, shared look between interlocutors, are sufficient to elicit targetlike responses. In contrast, longer eye gaze might reflect different underlying processes, perhaps indicating that the L2 speaker has misinterpreted the RA's recasts or is unable to reformulate.

To determine how L2 speakers' eye gaze of different durations was associated with targetlike responses, a *post hoc* contingency analysis was conducted. L2 speaker eye gaze was converted to a nominal variable with four duration quartiles, from shortest to longest, which was then compared with targetlike response rates (recasts without any L2 speaker eye gaze were excluded). A chi-square test indicated that the relationship between gaze duration quartile and targetlike response was significant: χ^2 (3, 848) = 20.99, p = .001, Cramer's V = .16). For the first three quartiles (the shorter durations), observed targetlike response rates were higher than expected. However, the adjusted standardized residuals (± 4.6) located significance in the fourth quartile, i.e., the longest eye gaze durations, where the observed targetlike response rate was lower than expected (38 versus 65, respectively) while the observed no reformulation rate was higher than expected (199 versus 172, respectively). Thus, the post hoc analysis provides some evidence that longer L2 speaker eye gaze durations may reflect different underlying processes than brief shared looks.

While this interpretation L2 speakers' eye gaze behaviors is certainly plausible, the

contrasting findings reported across the initial study and the current replication regarding the role

of L2 speakers' eye gaze might have also stemmed from differences in the participant

characteristics across the two studies. For example, the speakers' eye gaze durations may have

been impacted by linguistic or cultural factors (i.e., the participants in the initial study were more

linguistically and culturally homogenous) or by speaker proficiency and length of residence (i.e.,

the participants in this study had higher proficiency in English and longer residence in Canada).

In essence, L2 speakers' eye gaze behaviors might reflect multiple factors, such as speakers'

cultural preferences, attitudes towards the interlocutor, or pragmalingusitic competence in the

target language. Thus, future research must clarify whether L2 speakers' self-initiated eye gaze

has an independent role to play in L2 interaction that is distinct from the role served by mutual

gaze.

Finally, these replication results confirmed the results of prior research that showed a

difference in response rates based on the linguistic error targeted in recasts, with recasts for

lexical errors eliciting more targetlike responses than recasts targeting grammatical errors (Lyster

1998; Mackey et al. 2000; Carpenter et al. 2006; Loewen & Philp 2006). This finding appears to

be robust, cutting across various contexts, elicitation instruments, and learner profiles. Unlike

prior research though, the current dataset failed to confirm the effect of recast length whereby

shorter recasts have been more frequently noticed (Philp 2003) or recognized as being corrective

(Egi 2007) and have elicited higher repair rates (Sheen 2006). In contrast, we found similar

targetlike response rates for both short and clausal recasts. At least one reason for this might be

the tendency for unscripted interaction to contain shorter recasts focused on lexical errors, with

clausal recasts associated with grammatical errors. It was impossible to disentangle this

association in the current dataset, so future studies with more tightly-controlled interactional

feedback are needed to identify the interaction between recast length and target errors.

In a recent narrative and systematic review of replication research in the field of L2

learning, Marsden and colleagues (Marsden, Morgan-Short, Thompson, & Abugaber 2018)

estimated that only one self-labelled replication is published for every 400 articles, which is a

low rate compared to psychology. Set against this backdrop, this systematic replication

contributes to our field's efforts to increase the generalizability of research findings about the

underexplored topic of eye gaze in L2 interaction. This study was a systematic replication (or in

Marsden et al.'s terminology, a conceptual replication) of our prior research, involving multiple

changes to the original design—the participants (larger sample, more diverse language

backgrounds, higher proficiency), the experimental setup (placement of webcams), the

interlocutors (bilingual French-English speakers), and the interactive tasks (only two tasks, with

one more academically oriented). All these changes were meant to increase the quality of

research by minimizing error and increasing generalizability. Although Marsden et al.'s

systematic review did not reveal a clear association between the number of changes introduced in

a replication study and its support of the initial findings, we are cognizant that the multiple

changes introduced here may have factored into the current results. Finally, as they pointed out,

shared authorship between the initial and replication studies is associated with confirmation of

the findings, which may be driven by a greater fidelity in materials and procedures. In

23

recognition of this limitation and in the spirit of inviting further replication work, we invite other

researcher to explore the issues targeted here by making our materials available using publicly

accessible databases.

Cited as:
McDonough, K., Trofimovich, P., Dao, P., & Abashidze, D. (2018). Eye gaze and L2 speakers'
responses to recasts: A systematic replication study of McDonough, Crowther, Kielstra and
Trofimovich (2015). *Language Teaching*, 1-15.

24

Note

[1] Because each L2 speaker and RA contributed more than one recast episode to the dataset, a

model including predictor variables for the participant and RA was calculated (Saito, 1999).

Although both variables were significant predictors, their inclusion in the model had little effect

on the odds ratio or significance of the other predictor variables. In other words, although the

data are not completely independent, the dependency is not significantly affecting the model.

Published online on Dec 18, 2018

25

Acknowledgements

<center>26</center>

References

Authors, in review

Allen, L.Q. (2000). Nonverbal accommodation in foreign language teacher talk. *Applied Language Learning* 11.1, 155–176.

Batstone, R. (2010). Issues and options in sociocognition. In R. Batstone (ed.), *Sociocognitive perspective on language use and language learning*. Oxford: Oxford University Press, 3–23.

Bavelas, J., Coates, L. & Johnson, T. (2002). Listener responses as collaborative process: The role of gaze. *Journal of Communication* 52.3, 566–580. https://doi.org/10.1111/j.1460-2466.2002.tb02562.x

Block, D. (2003). *The social turn in second language acquisition.* Washington, DC: Georgetown University Press.

Carpenter, H., Jeon, K. S., MacGregor, D. & Mackey, A. (2006). Learners' interpretation of recasts. *Studies in Second Language Acquisition*, *28*, 209–236. https://doi.org/10+10170S0272263106060104

Choi, H. & Iwashita, N. (2016). Interactional behaviours of low-proficiency learners in small group work. In M. Sato & S. Ballinger (eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda*. Amsterdam: John Benjamins, 113–134.

Conklin, K, & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second

language research. *Second Language Research* 32, 453–467.

https://doi.org/10.1177/0267658316637401

Davies, M. (2006). Paralinguistic focus on form. *TESOL Quarterly* 40.4, 841-855. doi:

10.2307/40264316

Dobao, A. (2014). Attention to form in collaborative writing tasks: Comparing pair and small

group interaction. *Canadian Modern Language Review* 70.1, 158–187.

https://doi.org/10.3138/cmlr.1768

Egi, T. (2007). Interpreting recasts as linguistic evidence: The roles of linguistic target, length,

and degree of change. *Studies in Second Language Acquisition* 29.4, 511–537.

https://doi.org/0+10170S0272263107070416

Faraco, M. & Kida, T. (2008). Gesture and the negotiation of meaning in a second language

classroom. In S. G. McCafferty & G. Stam (eds.), *Gesture: Second language acquisition

and classroom research.* London: Routledge, 280–297.

Firth, A. & Wagner, J. (1997). On discourse, communication and some fundamental concepts in

SLA research. *The Modern Language Journal* 81.3, 285–300.

https://doi.org/10.2307/329302

Floyd, S., Manrique, E., Rossi, G. & Torreira, F. (2016) Timing of visual bodily behavior in

repair sequences: Evidence from three languages. *Discourse Processes* 53.1, 175–204,

https://doi.org/10.1080/0163853X.2014.992680

28

Goo, J. (2016). Corrective feedback and working memory capacity: A replication. In G.

Granena, D. Jackson, & Y. Yilmaz (eds.), *Cognitive individual differences in second language processing and acquisition*. Amsterdam: John Benjamins, 279–302.

https://doi.org/10.1075/bpa.3.13goo

Goo, J. & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition* 35.1, 127–165. https://doi.org/10.1017/S0272263112000708

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York, NY: Academic Press.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition. *International Review of Applied Linguistics* 44.1, 103–124.

https://doi.org/10.1515/IRAL.2006.004

Gullberg, M. (2010). Methodological reflections on gesture analysis in second language acquisition and bilingualism research. *Second Language Research* 26.1, 75–102.

https://doi.org/10.1177/0267658309337639

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York, NY: Wiley.

Hulstijn, J., Young, R., Ortega, L., Bigelow, M., DeKeyser, R., Ellis, N., Lantolf, J. P., Mackey, A. & Talmy, S. (2014). Bridging the gap: cognitive and social approaches to research in second language learning and teaching. *Studies in Second Language Acquisition* 36.3, 361–421. https://doi.org/10.1017/S0272263114000035

Cited as:
McDonough, K., Trofimovich, P., Dao, P., & Abashidze, D. (2018). Eye gaze and L2 speakers' responses to recasts: A systematic replication study of McDonough, Crowther, Kielstra and Trofimovich (2015). *Language Teaching*, 1-15.

Lazaraton, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A

microanalytic inquiry. *Language Learning* 54.1, 79–117. https://doi.org/10.1111/j.1467-

9922.2004.00249.x

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language

Learning* 60.2, 309–365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

Loewen, S. & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics,

explicitness, and effectiveness. *The Modern Language Journal* 90.4, 536–556.

https://doi.org/10.1111/j.1540-4781.2006.00465.x

Long, M. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum.

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in

Second Language Acquisition* 20.1, 51–81. https://doi.org/10.1017/S027226319800103X

Lyster, R. & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback

research. *Studies in Second Language Acquisition* 35.1, 167–184.

https://doi.org/10.1017/S027226311200071X

Lyster, R. & Saito, K. (2010). Oral feedback in classroom SLA. *Studies in Second Language

Acquisition* 32.2, 265–302. https://doi.org/10.1017/S0272263109990520

Mackey, A. & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research

synthesis. In A. Mackey (ed.), *Conversational interaction in SLA: A collection of

empirical studies*. New York: Oxford University Press, 408–452.

Mackey, A., Gass, S. & McDonough, K. (2000). How do learners perceive interactional

feedback? *Studies in Second Language Acquisition* 22.4, 471–497.

Marsden, E., Morgan-Short, K., Thompson, S. & Abugaber, D. (2018). Replication in second

language research: Narrative and systematic reviews and recommendations for the field.

Language Learning. Published online 5 March 2018. https://doi.org/10.1111/lang.12286

McDonough, K., Crowther, D., Kielstra, P. & Trofimovich, P. (2015). Exploring the potential

role of eye gaze in eliciting English L2 speakers' responses to recasts. *Second Language*

Research 31.4, 563–575. https://doi.org/10.1177/0267658315589656

Miller, P.C. & Pan, W. (2012) Recasts in the L2 classroom: A meta-analytic review.

International Journal of Educational Research 56.1, 48–59.

https://doi.org/10.1016/j.ijer.2012.07.002

Moore, D. & Dunham, P. J. (1995). *Joint attention: Its origin and role in development.* Hillsdale,

NJ: Lawrence Erlbaum.

Nakatsukasa, K. (2016). Efficacy of recasts and gestures on the acquisition of locative

prepositions. *Studies in Second Language Acquisition* 38.4, 771–799

https://doi.org/10.1017/S0272263115000467

Philp, J. (2003). Constraints on "noticing the gap." *Studies in Second Language Acquisition* 25.1,

99–126. https://doi.org/10.1017/S0272263103000044

Philp, J. & Duchesne, S. (2008). When the gate opens: the interaction between social and

linguistic goals in child second language development. In J. Philp, R. Oliver, & A.

Mackey (eds.), *Second language acquisition and the younger learner: Child's play?*

Philadelphia: John Benjamins, 83–103.

31

Philp, J. & Mackey, A. (2010). Interaction research: What can socially informed approaches

offer to cognitivists (and vice versa)? In R. Batstone (ed.), *Sociocognitive perspectives on*

language use and language learning. Oxford: Oxford University Press, 210–224.

Plonsky, L. & Brown, D. (2015). Domain definition and search techniques in meta-analyses of

L2 research (Or why 18 meta-analyses of feedback have different results). *Second*

Language Research 31.2, 267–278. https://doi.org/10.1177/0267658314536436

Plonsky, L. & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The

case of interaction research. *Language Learning* 61.2, 325–66. doi: 10.1111/j.1467-

9922.2011.00640.x

Poteau, C. (2017). *Pedagogical innovations in foreign language learning via interlocutor*

familiarity. Cambridge: Cambridge Scholar Publishing.

R Core Team (2016). R: A language and environment for statistical computing (Version 3.5.0)

[Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved

17 September 2017 from https://www.R-project.org

Rassaei, E. (2015). The effects of foreign language anxiety on EFL learners' perceptions of oral

corrective feedback. *Innovation in Language Learning and Teaching* 9.1, 87–101.

https://doi.org/10.1080/17501229.2013.837912

Revesz, A. (2012). Working memory and the observed effectiveness of recasts on different L2

outcome measures. *Language Learning* 62.1, 93–132. https://doi.org/10.1111/j.1467-

9922.2011.00690.x

Richardson, D. C. & Dale, R. (2005). Looking to understand: The coupling between speakers'

32

and listeners' eye movements and its relationship to discourse comprehension. *Cognitive*

   Science 29.6, 1045–1060. https://doi.org/10.1207/s15516709cog0000_29

Rossano, F., Brown, P. & Levinson, S. C. (2009) Gaze, questioning and culture. In J Sidnell

   (ed.), *Conversation analysis: Comparative perspectives.* Cambridge, UK: Cambridge

   University Press, 187–249.

Russell, J. & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2

   grammar: A meta-analysis of the research. In J. Norris and L. Ortega (eds.), *Synthesizing*

   research on language learning and teaching. Philadelphia, PA: John Benjamins, 133–

   164.

Saito, H. (1999). Dependence and interaction in frequency data analysis in SLA research. *Studies*

   in Second Language Acquisition 21.3, 453–475.

   https://doi.org/10.1017/S0272263199003046

Sato, M. & Viveros, P. (2016). Interaction or collaboration? Group dynamics in the foreign

   language classroom. In M. Sato & S. Ballinger (eds.), *Peer interaction and second*

   language learning: Pedagogical potential and research agenda. Amsterdam: John

   Benjamins, 91–112.

Saville-Troike, M. & Barto, K. (2017). *Introducing second language acquisition* (3rd ed.).

   Cambridge University Press.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner

   uptake. *Language Teaching Research* 10.4, 361–392. doi: 10.1191/1362168806lr203oa

Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language*

33

Learning 58.4, 835–874. https://doi.org/10.1111/j.1467-9922.2008.00480.x

Smotrova, T. & Lantolf, J. P. (2013). The function of gesture in lexically focused L2

instructional conversations. *Modern Language Journal* 97.2, 397–416.

https://doi.org/10.1111/j.1540-4781.2013.12008.x

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning* 52.1, 119–158.

https://doi.org/10.1111/1467-9922.00179

Storch, N. & Aldosari, A. (2012). Pairing learners in pair work activity. *Language Teaching*

Research 17.1, 31–48. https://doi.org/10.1177/1362168812457530

Trofimovich, P., Ammar, A. & Gatbonton, E. (2007). How effective are recasts? The role of

attention, memory, and analytical ability. In Mackey, A. (ed.), *Conversational interaction*

in second language acquisition: A collection of empirical studies. Oxford: Oxford

University Press, 171–195.

Young, A. & Tedick, D. (2016). Collaborative dialogue in a two-way Spanish/English

immersion classroom: Does heterogeneous grouping promoting peer linguistic

scaffolding? In M. Sato & S. Ballinger (eds.), *Peer interaction and second language*

learning: Pedagogical potential and research agenda. Amsterdam: John Benjamins,

135–162.

Wang, W. & Loewen, S. (2016). Nonverbal behavior and corrective feedback in nine ESL

university-level classrooms. *Language Teaching Research* 20.4, 459-478.

https://doi.org/0.1177/1362168815577239

Published online on Dec 18, 2018

Cited as:
McDonough, K., Trofimovich, P., Dao, P., & Abashidze, D. (2018). Eye gaze and L2 speakers'
responses to recasts: A systematic replication study of McDonough, Crowther, Kielstra and
Trofimovich (2015). *Language Teaching*, 1-15.

34

Ziegler, N. (2018). The contingency of recasts, learners' noticing, and L2 development: Insights

on saliency from multiple modalities. In S. Gass, P. Spinner, & J. Behney (eds.). *Salience

and SLA*. New York, NY: Routledge, 269–290.