



---

## Studies in Second Language Learning and Teaching

Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz

SSLT 8 (4). 2018. 721-754

<http://dx.doi.org/10.14746/sslit.2018.8.4.2>

<http://pressto.amu.edu.pl/index.php/sslit>

---

### *The L2 motivational self system: A meta-analysis*

Ali H. Al-Hoorie

Jubail Industrial College, Saudi Arabia

[hoorie\\_a@jic.edu.sa](mailto:hoorie_a@jic.edu.sa)

#### Abstract

This article reports the first meta-analysis of the L2 motivational self system (Dörnyei, 2005, 2009). A total of 32 research reports, involving 39 unique samples and 32,078 language learners, were meta-analyzed. The results showed that the three components of the L2 motivational self system (the ideal L2 self, the ought-to L2 self, and the L2 learning experience) were significant predictors of subjective intended effort ( $r_s = .61, .38, \text{ and } .41$ , respectively), though weaker predictors of objective measures of achievement ( $r_s = .20, -.05, \text{ and } .17$ ). Substantial heterogeneity was also observed in most of these correlations. The results also suggest that the strong correlation between the L2 learning experience and intended effort reported in the literature is, due to substantial wording overlap, partly an artifact of lack of discriminant validity between these two scales. Implications of these results and directions for future research are discussed.

*Keywords:* ideal L2 self; ought-to L2 self; L2 learning experience; L2 motivational self system; self-guides

#### 1. Introduction

In 2005, Dörnyei introduced the L2 motivational self system (L2MSS) as an attempt to explain individual differences in language learning motivation. The L2MSS is influenced by a number of theories, most notably possible selves theory (Markus & Nurius, 1986), self-discrepancy theory (Higgins, 1987), and the socio-educational

model (Gardner, 1979, 1985, 2010). A fundamental assumption in the L2MSS is that when the learner perceives a discrepancy between their current state and their future self-guide (i.e., ideal or ought), this discrepancy may function as a motivator to bridge the perceived gap and reach the desired end-state. In 2009, the first anthology testing this model appeared (Dörnyei & Ushioda, 2009b) reporting a number of empirical investigations that, according to Dörnyei (2009), “found solid confirmation for the proposed self system” (p. 31).

Subsequently, interest in this model increased exponentially in the language motivation field. Within just one decade, the L2MSS generated “an exceptional wave of interest with literally hundreds of studies appearing worldwide” (Dörnyei & Ryan, 2015, p. 91). In fact, in their comprehensive survey of over 400 recent publications, Boo, Dörnyei, and Ryan (2015) report that the L2MSS is currently the dominant theoretical framework in the field. Boo et al. (2015) attribute this dominance to the versatility of the model and its ability to accommodate a wide range of perspectives from different theoretical orientations.

The L2MSS consists of three main components (Dörnyei, 2005, 2009): the ideal L2 self, the ought-to L2 self, and the L2 learning experience. The ideal L2 self refers to the state one would ideally like to reach, thus representing one’s own hopes and wishes. The ought-to L2 self, on the other hand, refers to the state that others would want one to reach, thus representing the expectations projected by significant others. On a different level, the L2 learning experience concerns one’s experience in the immediate learning environment, involving aspects such as the teacher, the curriculum, and peers. The next section reviews the evidence each of these three components has generated.

## 2. Components of the L2MSS

### 2.1. The ideal L2 self

The ideal L2 self has received a significant amount of attention in recent literature. However, the results seem to have led to a range of conclusions in the field, some of which seem polarized. On the one hand, the predictive validity of the ideal L2 self has been described as “straightforward” (Dörnyei & Ushioda, 2011, p. 87), and as providing “solid confirmation” (Dörnyei, 2009, p. 31) in that “the emerging picture consistently supports [its] validity” (Dörnyei, 2014, p. 521). Similarly, Dörnyei and Ryan (2015) argue that “virtually all the validation studies reported in the literature found the L2 Motivation Self System providing a good fit for the data” (p. 91). Ghanizadeh and Rostami (2015) further state that the “results conclusively verified the model in virtually every context.” These comments generally refer to the ideal L2 self specifically (see also Ghanizadeh, Eishabadi, &

Rostami, 2016, p. 15; Henry & Cliffordson, 2015, p. 20; Islam, Lamb, & Chambers, 2013, p. 238; Teimouri, 2017, p. 683).

On the other hand, some other researchers expressed some reservation. For example, in their investigation of Korean secondary school students, Kim and Kim (2011) report that the ideal L2 self could not predict school grades. The researchers note that “being motivated by developing a vivid ideal L2 self through a dominant visual preference seems to be irrelevant to the level of academic achievement” (p. 36). Similarly, Lamb (2012) administered a C-test to Indonesian learners and found, again, that the ideal self could not predict proficiency. He therefore argued that although his participants “would like to see themselves as future users of English (ideal L2 self), what makes them more likely to invest effort in learning is whether they feel positive about the process of learning” (p. 1014). In the Canadian context, MacIntyre and Serroul (2015) examined the relationship between the ideal L2 self and actual L2 performance in their idiodynamic paradigm, which measures individual motivational variability on a per-second timescale. The researchers found “no evidence” (p. 126) that the ideal L2 self is dynamic or adapting to the changing task demands. In the Iranian context, Papi and Abdollahzadeh (2012) also found that the ideal L2 self does not predict actual classroom behavior. The researchers explain that:

the learners’ ideal image of their future self does not have much impact on their motivated behavior in English language classrooms or vice versa; that is, regardless of how well-developed the students’ ideal L2 self is, their actual motivated behavior in classroom activities will remain unaffected, and regardless of how motivated the students are in class, their ideal L2 selves will remain unchanged. (Papi & Abdollahzadeh, 2012, p. 588)

In the Saudi context, Moskovsky, Assulaimani, Racheva, and Harkins (2016) found the ideal L2 self to be a *negative* predictor of language proficiency. The researchers argue that, overall, the results “at best indicate a tenuous link between the self guides and achievement” (p. 650).

Thus, the emerging literature points to a rather complex picture. This could plausibly be due to certain factors, such as applicability of the model to different contexts or participants, or the use of different outcome measures. As explained in more detail below, a meta-analysis can help shed more light on such conflicting results.

## 2.2. The ought-to L2 self

In contrast to the controversy surrounding the ideal L2 self, there seems to be more agreement that the ought-to L2 self could benefit from some improvement. For example, Dörnyei and Chan (2013) acknowledge that “while [ought-to

selves] do play a role in shaping the learners' motivational mindset, in many language contexts they lack the energising force to make a difference in actual motivated learner behaviours by themselves" (p. 454). They then go on to explain that "while the participants perceived the external pressures on them as being valid and did *intend* to adjust their behavior accordingly, this intended effort was not manifested in their actual grades" (p. 454, original emphasis).

In recognition of the wanting nature of the ought-to L2 self construct, a number of developments have been proposed. Most of these developments argue for the need to incorporate the distinction between *own* and *other* standpoints in both the ideal and ought-to L2 selves. From this perspective, the ideal L2 self should be separated into two constructs, one representing one's own hopes and one significant others' hopes. Similarly, the ought-to L2 self should be bifurcated into obligations one would like to perform and obligations others expect one to perform (see Papi, Bondarenko, Mansouri, Feng, & Jiang, in press; Taylor, 2013).

For example, Thompson and Vásquez (2015) conducted a narrative study on three language teachers and argued that their data indicate a distinction between an ought-to L2 self and an anti-ought-to L2 self, the latter referring to one's own desires that are at odds with what the others expect from the individual. Lanvers (2016) conducted another qualitative study on language learners and argued that the ought–other standpoint should feature more prominently in educational contexts, as parents and teachers typically exert a lot of influence on students.

In one of the few quantitative studies testing the relevance of own–other standpoints to the language learning context, Teimouri (2017) developed questionnaire scales to measure each of the four proposed constructs: the ideal–own, ideal–other, ought–own, and ought–other. Interestingly, Teimouri found support for the distinction between own and other in the case of the ought-to L2 self, but not the ideal L2 self. Teimouri argued that ideals are highly internalized, and consequently they may not be separable into those that relate to one's own versus others' ideals.

However, in order to be able to evaluate the contribution of these developments and the extent to which they have advanced the original construct, it is important to have a frame of reference. That is, without quantifying the predictive validity of the original ought-to L2 self, it may not be immediately apparent how much of an improvement an alternative variation of this construct is. A meta-analysis can offer a baseline against which the effectiveness of reformation attempts can be evaluated.

### 2.3. The L2 learning experience

This construct has been variously labeled as 'the L2 learning experience' and as 'attitudes toward language learning.' All these terms refer to the same construct

because of the considerable overlap in the scales used to measure them (cf. You, Dörnyei, & Csizér, 2016, pp. 96-97). The L2 learning experience operates on a different level from either the ideal L2 self or the ought-to L2 self. Unlike them, the L2 learning experience is concerned with attitudes and evaluations of the present learning environment rather than a future-oriented self-guide. However, due to the increasing interest in self-guides in recent years (cf. Boo et al., 2015), very little attention has been paid to this construct. For example, Dörnyei describes the L2 learning experience as the situated, executive motive (Dörnyei, 2009, p. 29) and as the causal dimension (Dörnyei, 2005, p. 106) of the model. Beyond that, very little work has been done to clarify the role of such executive motives or the mechanisms that underlie their causal effect, making it the least theorized construct in the L2MSS (Ushioda, 2011, p. 201). Despite that, the L2 learning experience has been described as the strongest predictor in the L2MSS (e.g., Lamb, 2012; Teimouri, 2017).

Interestingly, the vast majority of studies testing this construct in our field has been observational. The standard design involves administering a questionnaire scale to learners and then examining the relationship (e.g., using correlation, regression, or structural equation modeling) between scores from this scale and from other criterion measures. However, this approach is prone to confounds, thus risking obtaining spurious results that do not underlie a genuine causal relationship. Beleche and colleagues point out the need for caution in interpreting observational studies:

The positive association between grades and course evaluations may also reflect initial student ability and preferences, instructor grading leniency, or even a favorable meeting time, all of which may translate into higher grades and greater student satisfaction with the course, but not necessarily to greater learning. (Beleche, Fairris, & Marks, 2012, p. 709)

Other potential factors shown to confound course evaluations include the teacher's age, ethnicity, gender, and even clothes and attractiveness (for reviews, see Ottoboni, Boring, & Stark, 2016; Stark & Freishtat, 2014). In fact, results by Ambady and Rosenthal (1993) show that students, simply after watching a very brief silent video (less than 30 seconds), form impressions about their teachers and that these first impressions then predict end-of-course evaluations. The presence of all of these biases has led some researchers to cast serious doubt on the value of course evaluation, with some considering any attempt to statistically adjust for the many biases involved to be practically "impossible" (Ottoboni et al., 2016, p. 10).

When it comes to experimental research, a number of educational studies conducted in different parts of the world – including Italy (Braga, Paccagnella, & Pellizzari, 2014), France (Boring, 2015), and the United States (Arbuckle & Williams, 2003; Carrell & West, 2010; MacNell, Driscoll, & Hunt, 2015) – have demonstrated

that student satisfaction with the course is biased (based on objective measures). The results of these studies also cast doubt on any clear (positive) causal relationship between satisfaction with the course and achievement. In fact, some of them found a *negative* relationship between satisfaction and success in subsequent, more advanced courses. For example, results by Braga et al. (2014) show that “teachers who are more effective in promoting future performance receive worse evaluations from their students” (p. 81).

In the present study, an attempt is made to meta-analyze the relationship between the L2 learning experience and language learning outcomes. The results are then used as a springboard to discuss the implications of results from observational studies and compare them to those from experimental studies.

### 3. Need for meta-analysis

A rigorous evaluation of a theory requires a systematic review of its accumulating literature. When sufficient quantitative reports become available, their results may be synthesized in a meta-analysis. A meta-analysis typically aims to estimate the magnitude (and confidence intervals) of the reported effect sizes, while moving away from a dichotomous significant versus non-significant outcome. A meta-analysis can also be helpful in shedding light on conflicting results. That is, it is plausible that conflicting results might to some extent be explainable by certain characteristics of different studies, such as type of participants, research design, or instruments used. For example, the literature on the ideal L2 self has drawn from different measures to date. Some researchers used subjective self-reports (i.e., intended effort), while others used more objective criteria (e.g., school grades and other achievement tests). It is plausible that different measures lead to different results. When used to test such hypotheses, a meta-analysis can potentially contribute to resolving debates in the literature.

### 4. The present study

Despite the growing number of studies drawing from the L2MSS, no systematic meta-analysis has been conducted on this literature to date. Instead, previous researchers have so far engaged in head-counting, such as tallying the number of published studies (e.g., Boo et al., 2015); or in vote-counting, such as describing the results of these studies as either supporting the theory or as ‘mixed’ (e.g., Dörnyei & Chan, 2013). Describing findings as mixed does not inform the reader about their average estimate, the width of its confidence interval, and whether any heterogeneity (i.e., variability of the estimate) found can or cannot be explained by moderators. Because a meta-analysis can address these questions, the

present study aimed to meta-analyze studies drawing from the L2MSS. More specifically, the primary research question guiding this meta-analysis is as follows:

RQ. What is the correlation between each of the three components of the L2MSS (the ideal L2 self, the ought-to L2 self, and the L2 learning experience) and educational outcomes (subjective and objective measures)?

This research question indicates a total of six correlations to be investigated: three correlations with subjective measures and three with objective measures. Categorizing outcome measures into subjective and objective was a rather pragmatic decision due to, as is explained in more detail below, the scarcity of studies utilizing objective measures in the field. The vast majority of studies in recent literature have used intended effort as their primary criterion variable. However, objective measures of actual language learning and achievement (e.g., grades and other standardized tests) represent an indispensable part of the overall picture. For example, Roth et al. (2015) argue that “school grades are crucial for accessing further scholastic and occupational qualification, and therefore, have an enormous influence on an individual’s life” (p. 118). Similarly, Moskovsky et al. (2016) claim that “therein lies the real test for the theory – in the capacity of the self guides to predict L2 achievement” (p. 643; see also Dörnyei & Chan, 2013, p. 454; Dörnyei & Ryan, 2015, p. 101). Indeed, language proficiency and achievement are an essential consideration for many stakeholders such as parents, teachers, and future employers.

Still, arguing that objective measures are ‘the real test’ of a theory might imply downplaying subjective measures, when in fact subjective measures might plausibly capture a dimension not captured by objective criteria. For completeness, therefore, the correlation between the two outcome measures was investigated to find out the degree of correspondence between them.

## 5. Method

### 5.1. Inclusion criteria

In order to be eligible for inclusion in this meta-analysis, the report must satisfy the following criteria:

1. It must involve a quantitative component. Qualitative and conceptual articles were excluded.
2. It must be about language learners. Reports about language teachers were excluded.
3. It must include at least one of the three components of the L2MSS.

4. It must include at least one outcome variable, such as school grades, objective tests, or subjective intended effort.
5. It must report the zero-order correlation between at least one component of the L2MSS and one outcome measure, or provide sufficient information to calculate it. Studies with only regression coefficients were excluded.
6. It must be published in English.
7. It must have been available by the start of June 2017.

## 5.2. Literature search

The literature search commenced with the article pool compiled by Boo et al. (2015), spanning the period from 2005 to 2014 ( $k = 283$ , excluding book chapters). To complement this list and to find more recent reports, a search was conducted in databases relevant to our field: ERIC, LLBA, MLA, ProQuest, and PsychINFO using the following keywords: ideal L2 self, ought-to L2 self, L2 learning experience, and L2 motivational self system. This resulted in a number of additional journal articles and unpublished theses ( $k = 51$ ). The list was then complemented by a Google Scholar search and by an ancestry search to ensure saturation ( $k = 21$ ). Furthermore, 19 edited volumes published since 2005 were inspected ( $k = 309$  chapters). Finally, a call for papers was announced at various relevant mailing lists, including BAALmail, Linguist List, myTESOL Lounge, Korea TESOL, and IATEFL Research SIG, as well as social media – resulting in further reports ( $k = 14$ ).

This search procedure has therefore resulted in a pool of 678 journal articles, book chapters, and unpublished manuscripts, ranging from conceptual to empirical, quantitative and qualitative, as well as duplicates (e.g., theses that were later turned to one, or more, publications). This pool of reports was subsequently examined against the inclusion criteria listed above. Eventually, 32 reports involving 39 unique samples and 32,078 language learners met all inclusion criteria. The lists of the included studies and of their characteristics are available in Appendices A and B.

## 5.3. Data analysis

*Software.* Comprehensive Meta Analysis 3.3 (Borenstein, Hedges, Higgins, & Rothstein, 2014) was used for all analyses. A random-effects model was implemented, since there was no reason to assume that all studies share one common effect size. Heterogeneity was examined using the  $I^2$ -statistic and its associated significance value. The presence of significant heterogeneity implies that the effect is highly variable and could potentially be explained by certain characteristics of different studies.



*Publication bias.* Publication bias refers to the situation where the outcome of a study has an effect on whether that study is eventually published. Studies reporting statistically significant results tend to be perceived as more interesting than those reporting non-significant results, and therefore the latter may not successfully complete the long and laborious publication process. The authors themselves can also become discouraged or lose interest, and consequently decide not to undergo the publication process. In some cases, the authors may believe that there must have been a mistake, especially when their results are not in line with mainstream views. This can lead to what is known as the file drawer problem (Rosenthal, 1979).

Publication bias may be inferred when small-scale studies, with statistically lower precision, report extreme values relative to larger-scale studies. Due to their lower power, some small studies are expected to find non-significant results simply by chance. However, when such small studies report significant results consistently, the likelihood that the literature is significant-biased increases. In the present meta-analysis, publication bias was examined using the Trim and Fill method (Duval & Tweedie, 2000a, 2000b). The Trim and Fill method is currently the most popular corrective technique to adjust for publication bias in contemporary meta-analytic literature (Simonsohn, Nelson, & Simmons, 2014).

*Inclusion criteria.* Initially, a second coder analyzed 10% of the reports independently against the inclusion criteria described above (Cohen's  $\kappa = .76$ ,  $p < .001$ ). Subsequently, discrepancies were resolved by discussion until 100% agreement was reached. Very few studies reported longitudinal investigations ( $k = 1$ ). In this case, the first time point was included. Also very few studies reported two measures for the L2 learning experience ( $k = 1$ ) or intended effort ( $k = 1$ ). In these cases, the two measures were averaged before inclusion in the analysis.

Most studies adopted the standard research design of administering questionnaire scales adopted with minor variations from Taguchi, Magid, and Papi (2009), typically translated to the participants' L1. Some reports were excluded for not reporting the results for Pearson correlation, such as instead reporting regression coefficients ( $k = 13$ ), the path coefficients in structural equation models ( $k = 11$ ), or other procedures ( $k = 2$ ). However, over 90% of these reports used intended effort as their criterion measure. Due to the relatively large number of reports drawing from intended effort that are already eligible for inclusion in the present meta-analysis, the excluded reports would have probably had a minor impact had they been included. This issue is discussed further in the Limitations section below (see Appendix C for a list of studies excluded for incomplete reporting).

*Published versus unpublished reports.* Unpublished reports are typically included in meta-analyses (Norris & Ortega, 2006). Although unpublished studies raise quality concerns, they may also represent studies with null results or with

results going against mainstream views – making them harder to publish. Other reports may have been completed as part of a degree program (e.g., MA or PhD) and publication was not subsequently pursued.

In the present meta-analysis, there were a number of unpublished reports ( $k = 6$ ). As a quality control procedure, moderation analysis was conducted to compare the results obtained from published and unpublished reports. The results showed that all comparisons were non-significant at the .05 level, thus providing no evidence that this small sample of unpublished reports have biased the results.

*Study quality.* Study quality is a perennial problem in meta-analysis, since low quality studies could potentially bias the results. While some researchers advocate excluding low quality studies altogether, others recommend including them and then conducting sensitivity analysis (e.g., Norris & Ortega, 2006). This is partly because study quality is not a straightforward concept, and different researchers may evaluate quality differently. Sensitivity analysis, however, can show whether the overall results are robust or highly influenced by the presence of studies with debatable quality.

In the present meta-analysis, the target statistic was Pearson correlation. Because this is a relatively straightforward procedure, it was expected that most reports would exhibit satisfactory quality. Following guidelines outlined by Dörnyei (2010), particular attention was also paid to psychometrics, such as using multi-item scales, providing suitable response options, and reporting reliability. All reports satisfying the inclusion criteria were analyzed by two coders independently (Cohen's  $\kappa = .87, p < .001$ ). Discussion of the minor discrepancies obtained led to the conclusion that a small number of reports ( $k = 2$ ) might potentially bias the results as the reliability of individual scales was missing. Sensitivity analysis was therefore conducted to examine the effect of excluding these two reports (see Results below).

*Subjective versus objective outcomes.* In the present sample, a large number of studies used intended effort as their criterion variable. In fact, even subjective self-ratings of proficiency can hardly be found in the literature. A smaller number of studies used more objective measures, including school grades and proficiency tests. Moderation analysis was conducted to compare the results obtained from school grades and from other objective measures. All tests were non-significant, thus justifying combining grades and objective measures into one category (called “achievement” henceforth).<sup>1</sup>

*Further moderators.* Unfortunately, it was not possible to test the moderating effect of some important learner characteristics, including age, gender,

---

<sup>1</sup> It has to be clarified that the term “subjective” does not mean less valid or less reliable. It simply means that it relies on the learner’s own perspective rather than on the results of a formal language test. Objective and subjective measures, therefore, serve different purposes.

and context. In terms of age, not a single study involving pre-secondary learners qualified for the final analysis, supporting Boo et al.'s (2015) observation that there is a "virtual absence" (p. 156) of research on younger learners in recent years. A few studies reported results for secondary and university learners separately, but the literature does not seem mature enough to meta-analyze the role of this variable since it was not always clear whether the target language was learned as part of a major or elective course or as an L2 or L3. In terms of gender, most studies reported the results combined for males and females, thus precluding any comparisons between the two genders. In terms of context, most investigations were conducted in a foreign language context, and only a small minority were in a second language context ( $k = 5$ , 3 of which were unpublished dissertations). Finally, a very small number of studies investigated a language other than English ( $k = 3$ ), supporting Dörnyei and Al-Hoorie's (2017) argument that the language motivation field is currently English-biased. Implications of these trends are discussed later.

## 6. Results

Table 1 reports the correlations between each of the three components of the L2MSS and the two outcome measures, as well as those between the two outcome measures themselves. In all cases, a sizable number of learners were included, with the smallest total being over 1,300. It is further evident from Table 1 that considerably fewer studies included a measure of actual achievement, while most used intended effort as their primary outcome variable.

Table 1 Correlations between the three L2MSS components and the two outcome variables

	<i>k</i> <i>N</i>		Intended effort				Achievement					
			<i>r</i>	95% CI		<i>p</i> <sup>2</sup>	<i>k</i>	<i>N</i>	<i>r</i>	95% CI		<i>p</i> <sup>2</sup>
				Lower	Upper					Lower	Upper	
Ideal L2 self	32	30,572	.611	.562	.655	97.21%	13	3,551	.202	.084	.315	90.76%
Sensitivity			—	—	—	—			.170	.046	.289	91.15%
Corrected			.611	.562	.655	97.21%			.103 <sup>ns</sup>	-.013	.218	93.70%
Ought-to L2 self	19	18,542	.379	.315	.440	94.21%	10	2,452	-.048 <sup>ns</sup>	-.107	.011	41.29% <sup>ns</sup>
Sensitivity			—	—	—	—			-.040 <sup>ns</sup>	-.107	.027	51.88%
Corrected			.379	.315	.440	94.21%			-.048 <sup>ns</sup>	-.107	.011	41.29% <sup>ns</sup>
L2 learning exp	18	19,586	.656	.590	.712	97.71%	7	1,369	.174	.026	.315	85.95%
Sensitivity			—	—	—	—			.137 <sup>ns</sup>	-.040	.306	89.35%
Corrected			.656	.590	.712	97.71%			.111 <sup>ns</sup>	-.029	.247	89.07%
Achievement	7	2,016	.116 <sup>ns</sup>	-.121	.341	96.02%						
Sensitivity			—	—	—	—						
Corrected			.116 <sup>ns</sup>	-.121	.341	96.02%						

Note. Exp = experience, ns = non-significant. Sensitivity analysis excluded two reports ( $n = 171$  total).

The three components of the L2MSS had positive correlations with intended effort, but dropped with achievement. There was also no overlap in the confidence intervals of each component's correlations with intended effort and with achievement, indicating that the coefficients are significantly different from each other. These findings might be used to explain some conflicting results in the literature: Researchers who used subjective measures found stronger support for the L2MSS than those who used objective measures. Furthermore, the correlation between intended effort and achievement was weak and non-significant, indicating that these two outcome measures cannot be used interchangeably. A stark illustration of this is found in the ought-to L2 self, where its correlation with Intended effort was positive and moderate in magnitude (.38), but reversed its sign with achievement (-.05). These findings point to the need to diversify outcome measures in the L2 motivation field to obtain a more comprehensive picture, rather than relying exclusively on intended effort.

The  $R^2$  values in Table 1 indicated that there was a wide and significant heterogeneity in most correlations. That is, with the exception of the one between the ought-to L2 self and achievement (which is non-significant), all other correlations exhibited heterogeneity in excess of 85% and higher. Some confidence intervals were also somewhat large, especially for the correlations between achievement and each of the ideal L2 self and the L2 learning experience. Such heterogeneity is to be expected since these studies were conducted in different parts of the world by different researchers working independently rather than adhering strictly to certain research protocols. Potential moderators might help explain this heterogeneity in future meta-analytic research when a sufficient pool of studies becomes available.

When it comes to sensitivity analysis, the two reports that were excluded for not reporting scale reliability happened to involve correlations with achievement only. The results after excluding these two reports are found in the 'sensitivity' rows in Table 1. The three correlations with achievement exhibited a minor drop, with that of the L2 Learning experience becoming no longer significant. When it comes to publication bias, adjusted values are reported in the 'corrected' rows in the table. Two correlations dropped to non-significance due to publication bias correction: the correlation between achievement and each of the ideal L2 self and the L2 learning experience. These two cases had relatively low sample sizes, suggesting a larger sample of studies utilizing objective measures is needed to obtain a more robust finding. It may also potentially suggest that there are further reports that show non-significant results but that could not be uncovered by the literature search of this study, despite the relatively generous inclusion criteria adopted (by including unpublished reports and book chapters) and a call for papers circulated widely in the field. Figure 1 presents a visual illustration of publication bias in the case of the ideal L2 self.

One surprising finding in Table 1 is the unusually high correlations of intended effort with the L2 learning experience. According to Dörnyei (2007), "if two tests correlate with each other in the order of 0.6, we can say that they measure more or less the same thing" (p. 223). While this may not be a hard-and-fast rule, the high correlations in Table 1 do raise discriminant validity concerns. This part of the analysis was therefore rerun to compare studies that examined the factorial structure of their scales (whether using classical test theory or item response theory) with those that did not.

The results in Table 2 indeed provide evidence that the high correlation between the L2 learning experience and intended effort may be a methodological artifact of not applying a factor-analytic procedure. The correlation between these two variables showed a significant drop from .68 to .41. A cursory look at the items used in studies that did not examine the factorial structure of their scales also showed considerable overlap. For example, one report used these two items: "Learning English is one of the most important aspects in my life" and "It is extremely important for me to learn English." Despite the close similarity of these two items, the former was used to measure attitudes toward learning English while the former intended effort. It is highly unlikely that these two items belong to two different latent variables. Unsurprisingly, that study reported a correlation of .91 between them for university majors, indicating that it may not be meaningful to distinguish between these two scales.

Table 2 Correlations between the three L2MSS components and intended effort for studies that applied a factor-analytic procedure and studies that did not

	<i>k</i>	<i>N</i>	<i>r</i>	95% CI		Intended effort	
				Lower	Upper	<i>Q</i>	<i>p</i>
<b>Ideal L2 self</b>							
With factor analysis	10	10,053	.548	.447	.636	2.695	.101
Without factor analysis	22	18,640	.637	.579	.689		
<b>Ought-to L2 self</b>							
With factor analysis	3	2,369	.378	.205	.528	< 0.001	.997
Without factor analysis	16	14,294	.378	.302	.449		
<b>L2 learning exp</b>							
With factor analysis	2	671	.408	.135	.624	6.051	.014
Without factor analysis	16	17,394	.680	.619	.733		

Note. A few studies reported ambiguous analyses (*k* = 2) and were therefore excluded.

Exp = experience, ns = non-significant

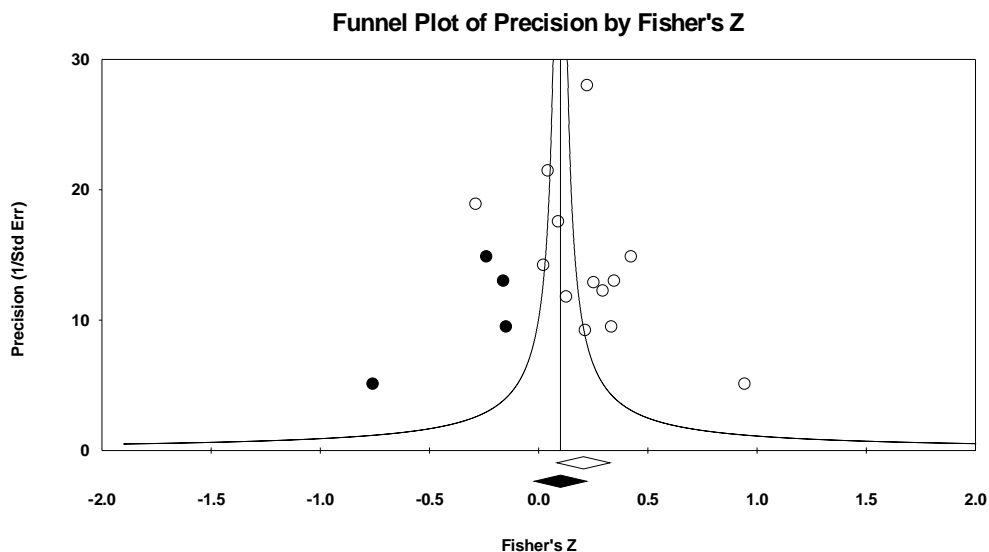


Figure 1 Funnel plot showing the relationship between the ideal L2 self and achievement

## 7. Discussion

The present meta-analysis has revealed a number of trends. One is that, perhaps for convenience, there is an abundance of research using intended effort as the primary criterion variable in recent language motivation research. On the other hand, there is a shortage of other outcome variables, resulting in an incomplete picture in the literature – especially since there was hardly any relationship between intended effort and other objective measures ( $r = .12$ ). Future research should attempt to draw from more diverse criterion measures in the hope of shedding more light on the multifaceted nature of motivation.

Another trend in recent literature is the lack of sufficient attention to important learner characteristics. More specifically, the present meta-analysis could not examine the effect of age, gender, or context. As for age, although older learners tend to be more accessible to researchers, it is possible that the dynamics of motivation is different at different ages (Kormos & Csizér, 2008). What motivates a 7-year-old might not motivate a 17-year-old (Nikolov, 1999). As for gender, it is often taken for granted that females exhibit higher motivation than males (You et al., 2016). However, systematic research to test this assumption is lacking, let alone attempting to explain it. As for context, the vast majority of recent motivation research has been conducted in foreign language contexts. This is in stark contrast to the social-psychological era, during which research in second language contexts was dominant (Al-Hoorie, 2017b). Hence, little is currently known

about the applicability of self-guides to second language contexts (see also Dörnyei & Ushioda, 2009a, pp. 352-353, for a similar argument).

A further trend is the dominance of English as the target language in recent research. English is indeed the global language and the most commonly taught nowadays. However, its global status may make the motivation to learn it distinct from the motivation to learn other languages (Dörnyei & Al-Hoorie, 2017). For example, a decision to learn a language like Danish or German typically needs to be accompanied by strong or personal reasons, especially when the aim is to achieve high proficiency. Learning English, in contrast, hardly needs a justification. This suggests a qualitative difference in the motivation to learn English versus the motivation to learn other languages. If this is the case, then the emphasis on English in recent literature risks deriving an incomplete theory of language learning motivation. This is an especially challenging task since the study of non-English languages is a rather complex subject. Non-English languages fall on different varieties such as minority, heritage, indigenous, and endangered languages, each with its unique set of contextual factors and conditions (Duff, 2017).

The following sections discuss the results of the present meta-analysis in relation to self-guides, the L2 learning experience, and intended effort. Limitations of this study are then highlighted.

### 7.1. Self-guides

In terms of the ideal L2 self, the results of the present meta-analysis showed that it correlated at .61 with intended effort and at .20 with achievement. In other words, the ideal L2 self accounts for around 37.2% of the variance in intended effort, but only about 4.1% in achievement. These results may help explain the conflicting findings in the literature: Studies relying on intended effort found strong support for the predictive validity of the ideal L2 self, while those drawing from other objective measures were less supportive.

Recently, Plonsky and Oswald (2014) have offered recommendations for field-specific benchmarks for interpreting the size of correlation coefficients: .25 small, .40 medium, and .60 large. If we follow these recommendations, the ideal L2 self is a strong predictor of intended effort, but approaching the small threshold in achievement. The relationship between the ideal L2 self and achievement is also smaller than the expected correlation between attitudes and behavior in social psychology ( $r = .38$ , Kraus, 1995). It is also smaller than the magnitude that aptitude ( $r = .49$ , Li, 2016) and intelligence ( $r = .54$ , Roth et al., 2015) explain in academic achievement, two established individual difference variables.

Given this modest magnitude, readers may wonder about the extent to which motivation contributes to language learning relative to the two classical

individual difference variables, intelligence and aptitude. Nevertheless, there seem to be a number of means to improve the predictive validity of the ideal L2 self when it comes to actual language achievement. First of all, the original conceptualization of the L2MSS comes with a set of conditions that, if not satisfied, self-guides are not expected to exhibit full power (Dörnyei, 2009). These conditions include the availability of the different self-guides, their vividness, plausibility, harmony, and activation, as well as having procedural strategies and being offset by a feared self. Although these conditions were proposed together with the inception of the theory itself, they have remained largely untested and hardly any attempts have been made to incorporate them into how self-guides are currently measured (Hessel, 2015).

Another potential direction is the incorporation of discrepancy. By definition, self-guides are not absolute constructs but relational to a future state. The hypothesized effect of the ideal L2 self, for example, resides in the discrepancy between a current state and a desired future state, not the future state per se. Unfortunately, this discrepancy is not currently featured in how self-guides are measured (Thorsen, Henry, & Cliffordson, 2017). The standard scale items used to measure the ideal L2 self are in the form of 'I can imagine myself...', which is admittedly ambiguous. As an illustration, if a learner cannot imagine herself mastering English someday, this could additionally mean that she does not believe she can do that (self-efficacy), that she does not want to do that (value of the activity), that she experiences a complete absence of motivation (amotivation), that she does not need to do that (e.g., she has already mastered English), or any other interpretations different learners might conjure up. Due to this ambiguity, it might be appropriate to relabel the standard ideal L2 self scale to *the imagined self*, and reserve the ideal L2 self label to an improved measure that can accommodate a current–future discrepancy that the L2MSS requires by definition.

A measure that can accommodate a current–future discrepancy does not have to be a close-ended questionnaire scale. In fact, self-discrepancy is not conceptualized as a conscious construct that the individual can readily self-report (Higgins, 1987). For this reason, Higgins (1987) criticized a study by Hoge and McCarthy (1983) for using experimenter-selected attributes and asking their participants about their discrepancies directly, describing this type of measure as *nonideographic*. An ideographic measure, in contrast, requires that the participant is the one who supplies attributes related to, say, their actual self and their ideal self separately. It is then the researcher's job to code these attributes in order to determine 'matches' and 'mismatches' between actual and ideal selves. The results may show that one participant has a large number of matches (i.e., little discrepancy), another with mostly mismatches (much discrepancy), and yet another with neither matches or mismatches (no relevance of discrepancy).



This approach has not been utilized in the language motivation field to date. Another approach that does not rely on close-ended questionnaires draws from reaction-time measures (e.g., Higgins, Shah, & Friedman, 1997; Shah & Higgins, 2001). The premise behind this approach is that higher accessibility leads to more efficient approach and avoidance tendencies unconsciously. Our field is yet to exploit the full potential of reaction-time measures to study unconscious aspects of motivation (Al-Hoorie, 2016a, 2016b, in press).

In terms of the ought-to L2 self, its predictive validity was markedly lower than that of the ideal L2 self in relation to both intended effort and achievement. As explained above, the wanting nature of the ought-to L2 self has already been pointed out by a number of scholars who recommended improvements. However, instead of leaving this construct behind in favor of newer constructs, it would be useful to attempt to understand why its theoretically anticipated effect has not been borne out.

One possible explanation is that the ought-to L2 self is – by definition (see Dörnyei, 2009, p. 29) – concerned only with the less internalized forms of motives. It pertains to someone else's expectations, rather than one's own ideals, and primarily functions in a preventive fashion. That is, since ought self-guides represent "minimal goals" (Higgins, 1998, p. 5) that are "imposed" (Dörnyei, 2009, p. 32) by one's peers, parents and authoritative figures, then learners may simply aim to achieve the minimum required to satisfy another person's desires, rather than fulfilling them more thoroughly as one might do with one's own ideals. Such minimal goals are less likely to sustain engagement in learning and enthusiasm about it in the long run. A similar picture emerges from possible selves theory. Markus and Nurius (1986) actually downplay the role of others in the formation of one's own possible selves. In their words, "others' perceptions of an individual are unlikely to reflect or to take into account possible selves" (p. 964). Markus and Nurius then point out that, "when we perceive another person, or another perceives us, this aspect of perception, under most conditions, is simply not evident and typically there is little concern with it" (p. 964). A similar picture emerges, again, from self-determination theory (Deci & Ryan, 2002), in which the less internalized forms of extrinsic motivation seem to be associated negatively with L2 achievement, but the more internalized forms are associated positively with it (e.g., Wang, 2008). Indeed, Mackay (2014, p. 394) reports that some of her interviewees construed external pressures to learn the language as a *demotivating* factor. All of this points to the need to reconsider the original conceptualization of the ought-to L2 self construct as a motivational factor, an assumption held in the field for more than a decade. It might be more appropriately conceived of, at least in some contexts, as a demotivating variable instead.

Another possible explanation is that current measurement practice does not distinguish between own-other standpoints in self-guides (Lanvers, 2016;

Teimouri, 2017; Thompson & Vásquez, 2015). However, before treading this path, a number of conceptual issues need to be addressed. First, introducing standpoints may make the different self-guides less clear-cut. That is, where do we draw the line between an ideal-own and ought-own, and between ideal-other and ought-other (see Dörnyei, 2009, pp. 13-14, for a similar argument). Second, as Dörnyei and Ushioda (2009a, p. 352) point out, degrees of internalization are inherent to self-determination theory. When degrees of internalization are used to justify the different self-guides (e.g., ideal-own versus ideal-other), theorists need to consider in what respects this new formulation is more than self-determination theory cast in self terminology. This is a crucial consideration since it is desirable to avoid a situation where different researchers within one field deal with more or less the same phenomena but independently due to different terminology (Dörnyei & Ryan, 2015).

A further consideration pertains to the proliferation of 'selves' witnessed in the field today. Some scholars (MacIntyre & Mackinnon, 2007; MacIntyre, Mackinnon, & Clément, 2009) argue that these selves are hardly more than mere metaphors, risking unnecessary redundancy and conceptual clutter. For example, MacIntyre and Mackinnon (2007) list over 60 self-related constructs in psychology, leading them to argue that "the multitude of overlapping concepts in the literature on the self is more confusing than integrativeness ever could be" (MacIntyre et al., 2009, p. 54). Just like psychology, the language motivation field is witnessing more and more selves being introduced, including anti-ought-to, rebellious, imposed, bilingual, multilingual, private, public, possible, and probable selves, but without sufficient attention to their construct validity or their overlap. In fact, it has become fashionable to introduce a new construct and suffix it with a 'self' even when existing constructs seem to exist (e.g., *anti-ought-to self* versus *reactance*, and *feared L2 self* versus *fear of failure*). Adding a new dimension to an existing construct (e.g., L2 reactance) may be more appropriate than introducing yet another 'self'. As Albert Bandura cautions,

a theory cast in terms of multiple selves plunges one into deep philosophical waters. It requires a regress of selves to a presiding superordinate self that selects and manages the collection of possible selves to suit given purposes. Actually, there is only one self that can visualize different desired and undesired futures and select courses of action designed to attain cherished futures and escape feared ones. (Bandura, 1997, p. 26)

## 7.2. The L2 learning experience

As reviewed above, the L2 learning experience has been described as the strongest predictor of intended effort. However, the results of the present meta-analysis suggest that the high correlation between the L2 learning experience and

intended effort may partly be an artifact of not implementing a factor-analytic procedure. A cursory look at the literature suggests that the importance of examining the factorial structure of scales is not appreciated. Researchers, reviewers, and editors seem satisfied with a quick Cronbach analysis showing a reliability of around .70. However, reliability assumes that the scale is already unidimensional, and when it is not, reliability can be artificially inflated (see Al-Hoorie & Vitta, in press; Green, Lissitz, & Mulaik, 1977; Sijtsma, 2009). Based on the present results, it is recommended that researchers routinely present the results of a factor analytic procedure (whether from classical test theory or item response theory) to establish convergent and discriminant validity among their scales, along with their reliabilities.

In contrast to its correlation with intended effort, the L2 learning experience had a modest correlation with achievement ( $r = .17$ ). This suggests that, to date, the small number of studies that have examined the correlation between this variable and achievement do not support a strong association. Furthermore, little theoretical analysis is available to explain why this association should be causal in the first place (Ushioda, 2011), especially since virtually all studies included in the present meta-analysis were observational. Neither is this modest association totally inconsistent with experimental studies (on non-L2 learning) that do not support a causal relationship between student evaluation of the course and educational outcomes (Arbuckle & Williams, 2003; Boring, 2015; Braga et al., 2014; Carrell & West, 2010; MacNell et al., 2015). Having a positive attitude toward the course and its teacher may not necessarily imply better learning, even if the learner believes so. Indeed, it is not an unusual experience for a learner to get the 'impression' that they have mastered the subject, but to subsequently realize from a test that there were significant gaps in their knowledge. This misleading impression of mastery may be attributed to different reasons, including a teacher with a charismatic personality or simply an entertaining approach (see Al-Hoorie, 2017a, for a more detailed review).

Evidence of this misleading impression has been demonstrated graphically in a classic experiment titled 'The Doctor Fox Lecture: A paradigm of educational seduction' (Naftulin, Ware, & Donnelly, 1973). These researchers recruited a professional actor to give a lecture about game theory (a subject he knew nothing about). The actor was given a fake name, Dr. Myron L. Fox, and was introduced to the unsuspecting audience as an expert in the application of mathematics to human behavior. Drawing from his acting skills, the actor peppered his lecture with some humor as well as meaningless, conflicting, and irrelevant information. At the same time, he sounded authoritative and exhibited a charismatic personality. Despite the empty content of the lecture, the audience reported having enjoyed the lecture and even learned from it (in fact, one

person even reported that s/he had read some of the speaker's publications!). We can confidently argue that, despite this favorable impression, no learning or any knowledge transmission about game theory occurred during that lecture. 'Dr. Fox' simply did not know the material. The feeling of having learned from the lecture is little more than a misattribution. The audience simply enjoyed the charismatic and authoritative personality of lecturer, but then misattributed this enjoyment to the informativeness of the lecture. Naftulin et al. (1973) conclude that "student satisfaction with learning may represent little more than the illusion of having learned" (p. 630). This is now known as the Dr. Fox effect.<sup>2</sup>

These results point to the urgent need for experimental studies in the language motivation field for testing causal assumptions. One reason for the paucity of experiments has to do with the numerous practical and logistic considerations involved (see, e.g., Csizér & Magid, 2014, Part 4). Still, language motivation researchers could take their cues from other SLA areas where experiments are very common. When it comes to instructional effects, for example, Plonsky (2013) reports that experimental studies are about twice more common in classroom research than are observational studies. Experimental studies are also needed to examine pedagogical implications derived from observational studies (e.g., Dörnyei & Kubanyiova, 2014; Hadfield & Dörnyei, 2013). It is not unimaginable that some of these recommendations might turn out premature. If some implications do turn out to be premature, this could ultimately undermine the field's credibility.

### 7.3. Intended effort

In the L2MSS tradition, self-reported intended effort has been frequently called the criterion measure (sometimes with capital C and M). Although any outcome can be described as a criterion measure (since it simply refers to the dependent variable), the convention of calling intended effort as the criterion measure is nowadays seen everywhere in research reports – from scale descriptions, through results tables, to structural equation models. Another euphemism is 'motivated behavior'. In reality, however, this scale typically refers to intended effort rather than observation of actual behavior.

Still, a subjective measure is not necessarily less valid. The use of a subjective measure could provide unique insights that more objective measures might not capture. Nevertheless, there are at least two important considerations to take into account with regard to this scale. First, the items in this scale

---

<sup>2</sup> In the original experiment, Naftulin et al. videotaped the lecture by Dr. Fox. Some of its footage is now available on YouTube: [www.youtube.com/watch?v=RcxW6nrWwtc](http://www.youtube.com/watch?v=RcxW6nrWwtc) (accessed 28 September 2017).

tend to be generic, while generic intentions are less likely to translate into behavior (Fishbein & Ajzen, 2010). This is especially because conscious thought suffers from substantial blind spots when it comes to predicting how one will actually behave (see Al-Hoorie, 2015, 2016a, 2016b, 2017b, for a more detailed discussion). Following Wolters and Taylor (2012), intended effort could be made more specific by recognizing the different 'dimensions' of motivated behavior. In one dimension, some activities reflect *behavioral engagement* while others reflect *academic engagement*. Behavioral engagement includes class participation and other overt behavioral effort. Academic engagement, on the other hand, refers to time spent on learning tasks and amount of assignments completed. Although both constitute 'effort', the latter reflects more quality engagement. A second dimension of effort is whether it is *universal* and *optional*. Universal engagement refers to the activities that all students are expected to engage in, such as attending class and doing homework. In contrast, optional engagement refers to going beyond the expected of the typical student, by showing initiative and volunteering for relevant extracurricular activities. A third dimension is the need to consider engagement in *adaptive* versus *maladaptive* forms of behavior. A learner may engage in adaptive learning behaviors, but might at the same time also engage in other maladaptive behaviors (e.g., procrastination, defensive pessimism, and other forms of self-handicapping). Focusing on adaptive behaviors only might miss important pieces from the overall picture. A final dimension is the need to consider *agentive* versus *non-agentive* behavior. Effort expended by the learner that is planful and purposeful should count as more than the effort that merely reflects norm following. "Students coerced to finish worksheets using specific tactics rigidly dictated by a teacher may appear cognitively and behaviorally engaged" (Wolters & Taylor, 2012, p. 645) but not necessarily actually motivated. Adopting such level of specificity would likely enrich our perspective on learning motivation and open up interesting directions for future research.

Second, the use of intended effort leads a conceptual difficulty. Theoretical clarity requires observing "the motivation → behavior → outcome chain" (Dörnyei, 2005, p. 73) because "If we want to draw more meaningful inferences about the impact of various motives, it is more appropriate to use some sort of a *behavioural measure* as the criterion/dependent variable" (Dörnyei & Ushioda, 2011, p. 200, original emphasis). Intended effort does not seem to qualify as representative of the 'behavior' piece of the chain – until it is actually performed. This is why, in his review of the L2MSS, Gardner (2010) maintains that "they relate one measure based on verbal report to another measure based on verbal report" (p. 73). A theoretical justification for the use of intended effort as an outcome measure is needed to clarify what we can learn from this construct and in which contexts.

## 8. Limitations and conclusion

Because this is the first meta-analysis of the L2MSS, the present study inevitably has a number of limitations. The number of studies using a criterion measure other than intended effort is relatively small. This small number resulted in relatively wide confidence intervals, and further showed evidence of publication bias. This small sample also led to a pragmatic decision to combine all objective measures into one group. However, it is not implausible that different outcome measures would exhibit different results (e.g., end-of-year grades versus researcher-administered tests).

Therefore, the present meta-analysis must be considered a meta-analysis-in-progress and be updated once a sufficient number of studies using different outcome measures become available. Although a number of studies were excluded for not reporting correlation results, most of these studies followed the general trend of relying on intended effort rather than other outcome measures. The same applies to potential moderators, including age, gender, context, and target language (see also Ellis, 2006).

The results also draw attention to the urgent need for experimental studies in the language motivation field. For historical reasons, our field has relied heavily on observational questionnaire-based research designs. At the same time, many arguments in the field have causal implications, and even pedagogical recommendations for classroom teachers. In fact, making a list of pedagogical implications has become a default expectation from researchers (and graduate students), even when their research is observational. Without experimental research to support such pedagogical recommendations, this practice may be at best misleading, and at worst damaging to the field. However, overcoming the various logistics involved in conducting experimental research – whether inside or outside the classroom – would eventually lead to a science that is more instructive to classroom practice and to language learning in general.

### Acknowledgements

I would like to thank Robert Gardner, Zoltán Dörnyei, Richard Clement, Peter MacIntyre, Diane Larsen-Freeman, Kim Noels, Sarah Mercer, Martin Lamb, Alastair Henry, Saadat Saeed, Neil McClelland, Luke Plonsky, and two anonymous reviewers for their comments on an earlier draft. I would also like to thank Phil Hiver and Joe Vitta for their assistance in data coding.

References

- Al-Hoorie, A. H. (2015). Human agency: Does the beach ball have free will? In Z. Dörnyei, P. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 55-72). Bristol, UK: Multilingual Matters.
- Al-Hoorie, A. H. (2016a). Unconscious motivation. Part I: Implicit attitudes toward L2 speakers. *Studies in Second Language Learning and Teaching*, 6(3), 423-454. <https://doi.org/10.14746/sslit.2016.6.3.4>
- Al-Hoorie, A. H. (2016b). Unconscious motivation. Part II: Implicit attitudes and L2 achievement. *Studies in Second Language Learning and Teaching*, 6(4), 619-649. <https://doi.org/10.14746/sslit.2016.6.4.4>
- Al-Hoorie, A. H. (2017a). *Implicit attitudes in language learning*. Unpublished PhD thesis, University of Nottingham, UK.
- Al-Hoorie, A. H. (2017b). Sixty years of language motivation research: Looking back and looking forward. *SAGE Open*, 7(1). <https://doi.org/10.1177/2158244017701976>
- Al-Hoorie, A. H. (in press). Motivation and the unconscious. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave Macmillan handbook of motivation for language learning*. Basingstoke, UK: Palgrave Macmillan.
- Al-Hoorie, A. H., & Vitta, J. P. (in press). The seven sins of L2 research: A review of 30 journals' statistical quality and their CiteScore, SJR, SNIP, JCR Impact Factors. *Language Teaching Research*. <https://doi.org/10.1177/1362168818767191>
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64(3), 431-441. <https://doi.org/10.1037/0022-3514.64.3.431>
- Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles*, 49(9), 507-516. <https://doi.org/10.1023/A:1025832707002>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review*, 31(5), 709-719. <https://doi.org/10.1016/j.econedurev.2012.05.001>
- Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005-2014: Understanding a publication surge and a changing landscape. *System*, 55, 145-157. <https://doi.org/10.1016/j.system.2015.10.006>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2014). *Comprehensive Meta Analysis (Version 3.3)*. Englewood, NJ: Biostat.
- Boring, A. (2015). *Gender biases in student evaluations of teachers*. Document de travail OFCE. Paris, France. Retrieved from <https://www.ofce.fr/pdf/dtravail/WP2015-13.pdf>

- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71-88. <https://doi.org/10.1016/j.econedurev.2014.04.002>
- Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3), 409-432. <https://doi.org/10.1086/653808>
- Csizér, K., & Magid, M. (Eds.). (2014). *The impact of self-concept on language learning*. Bristol, UK: Multilingual Matters.
- Deci, E. L., & Ryan, R. M. (Eds.). (2002). *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. London, UK: Lawrence Erlbaum.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford, UK: Oxford University Press.
- Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9-42). Bristol, UK: Multilingual Matters.
- Dörnyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). New York: Routledge.
- Dörnyei, Z. (2014). Motivation in second language learning. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 518-531). Boston, MA: Cengage Learning.
- Dörnyei, Z., & Al-Hoorie, A. H. (2017). The motivational foundation of learning languages other than Global English. *Modern Language Journal*, 101(3), 455-468. <https://doi.org/10.1111/modl.12408>
- Dörnyei, Z., & Chan, L. (2013). Motivation and vision: An analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, 63(3), 437-462. <https://doi.org/10.1111/lang.12005>
- Dörnyei, Z., & Kubanyiova, M. (2014). *Motivating learners, motivating teachers: Building vision in the language classroom*. Cambridge, UK: Cambridge University Press.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge.
- Dörnyei, Z., & Ushioda, E. (2009a). Motivation, language identities and the L2 self: Future research directions. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 350-356). Bristol, UK: Multilingual Matters.
- Dörnyei, Z., & Ushioda, E. (Eds.). (2009b). *Motivation, language identity and the L2 self*. Bristol, UK: Multilingual Matters.
- Dörnyei, Z., & Ushioda, U. (2011). *Teaching and researching motivation* (2nd ed.). Harlow, UK: Pearson.



- Duff, P. A. (2017). Commentary: Motivation for learning languages other than English in an English-dominant world. *Modern Language Journal*, 101(3), 597-607. <https://doi.org/10.1111/modl.12416>
- Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89-98. <https://doi.org/10.1080/01621459.2000.10473905>
- Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Ellis, N. C. (2006). Meta-analysis, human cognition, and language learning. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 301-322). Amsterdam, The Netherlands: John Benjamins.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York: Psychology Press.
- Gardner, R. C. (1979). Social psychological aspects of second language acquisition. In H. Giles & R. N. St. Clair (Eds.), *Language and social psychology* (pp. 193-220). Oxford, UK: Blackwell.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London, UK: Edward Arnold.
- Gardner, R. C. (2010). *Motivation and second language acquisition: The socio-educational model*. New York: Peter Lang.
- Ghanizadeh, A., Eishabadi, N., & Rostami, S. (2016). Motivational dimension of willingness to communicate in L2: The impacts of criterion measure, ideal L2 self, family influence, and attitudes to L2 culture. *International Journal of Research Studies in Education*, 5(3), 13-24. <https://doi.org/10.5861/ijrse.2015.1261>
- Ghanizadeh, A., & Rostami, S. (2015). A Dörnyei-inspired study on second language motivation: A cross-comparison analysis in public and private contexts. *Psychological Studies*, 60(3), 292-301. <https://doi.org/10.1007/s12646-015-0328-4>
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37(4), 827-838. <https://doi.org/10.1177/001316447703700403>
- Hadfield, J., & Dörnyei, Z. (2013). *Motivating learning*. New York: Routledge.
- Henry, A., & Cliffordson, C. (2015). The impact of out-of-school factors on motivation to learn English: Self-discrepancies, beliefs, and experiences of self-authenticity. *Applied Linguistics*, Advance online publication. <https://doi.org/10.1093/applin/amv060>
- Hessel, G. (2015). From vision to action: Inquiring into the conditions for the motivational capacity of ideal second language selves. *System*, 52, 103-114. <https://doi.org/10.1016/j.system.2015.05.008>

- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94(3), 319-340. <https://doi.org/10.1037/0033-295X.94.3.319>
- Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In P. Z. Mark (Ed.), *Advances in Experimental Social Psychology* (Vol. 30, pp. 1-46). New York: Academic Press.
- Higgins, E. T., Shah, J., & Friedman, R. (1997). Emotional responses to goal attainment: Strength of regulatory focus as moderator. *Journal of Personality and Social Psychology*, 72(3), 515-525. <https://doi.org/10.1037/0022-3514.72.3.515>
- Hoge, D. R., & McCarthy, J. D. (1983). Issues of validity and reliability in the use of real-ideal discrepancy scores to measure self-regard. *Journal of Personality and Social Psychology*, 44(5), 1048-1055. <https://doi.org/10.1037/0022-3514.44.5.1048>
- Islam, M., Lamb, M., & Chambers, G. (2013). The L2 motivational self system and national interest: A Pakistani perspective. *System*, 41(2), 231-244. <https://doi.org/10.1016/j.system.2013.01.025>
- Kim, Y.-K., & Kim, T.-Y. (2011). The effect of Korean secondary school students' perceptual learning styles and ideal L2 self on motivated L2 behavior and English proficiency. *Korean Journal of English Language and Linguistics*, 11(1), 21-42.
- Kormos, J., & Csizér, K. (2008). Age-related differences in the motivation of learning English as a foreign language: Attitudes, selves, and motivated learning behavior. *Language Learning*, 58(2), 327-355. <https://doi.org/10.1111/j.1467-9922.2008.00443.x>
- Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, 21(1), 58-75. <https://doi.org/10.1177/0146167295211007>
- Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, 62(4), 997-1023. <https://doi.org/10.1111/j.1467-9922.2012.00719.x>
- Lanvers, U. (2016). Lots of selves, some rebellious: Developing the self discrepancy model for language learners. *System*, 60, 79-92. <https://doi.org/10.1016/j.system.2016.05.012>
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, 38(4), 801-842. <https://doi.org/10.1017/S027226311500042X>
- MacIntyre, P. D., & Mackinnon, S. P. (April, 2007). *From integrative motivation to possible selves: The baby, the bathwater, and the future of language learning motivation research*. Paper presented at the American Association for Applied Linguistics (AAAL), Costa Mesa, California, USA.
- MacIntyre, P. D., Mackinnon, S. P., & Clément, R. (2009). The baby, the bathwater, and the future of language learning motivation research. In Z. Dörnyei & E.

- Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 43-65). Bristol, UK: Multilingual Matters.
- MacIntyre, P. D., & Serroul, A. (2015). Motivation on a per-second timescale: Examining approach-avoidance motivation during L2 task performance. In Z. Dörnyei, P. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 109-138). Bristol, UK: Multilingual Matters.
- Mackay, J. (2014). Applications and implications of the L2 motivational self system in a Catalan EFL context. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 377-400). Bristol, UK: Multilingual Matters.
- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291-303. <https://doi.org/10.1007/s10755-014-9313-4>
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41(9), 954-969. <https://doi.org/10.1037/0003-066X.41.9.954>
- Moskovsky, C., Assulaimani, T., Racheva, S., & Harkins, J. (2016). The L2 motivational self system and L2 achievement: A study of Saudi EFL learners. *The Modern Language Journal*, 100(3), 641-654. <https://doi.org/10.1111/modl.12340>
- Naftulin, D. H., Ware, J. E., Jr., & Donnelly, F. A. (1973). The Doctor Fox Lecture: A paradigm of educational seduction. *Academic Medicine*, 48(7), 630-635.
- Nikolov, M. (1999). "Why do you learn English?" "Because the teacher is short." A study of Hungarian children's foreign language learning motivation. *Language Teaching Research*, 3(1), 33-56. <https://doi.org/10.1177/136216889900300103>
- Norris, J. M., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3-50). Amsterdam, The Netherlands: John Benjamins.
- Ottoboni, K., Boring, A., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1>
- Papi, M., & Abdollahzadeh, E. (2012). Teacher motivational practice, student motivation, and possible L2 selves: An examination in the Iranian EFL context. *Language Learning*, 62(2), 571-594. <https://doi.org/10.1111/j.1467-9922.2011.00632.x>
- Papi, M., Bondarenko, A., Mansouri, S., Feng, L., & Jiang, C. (in press). Rethinking L2 motivation research: The 2x2 model of L2 self-guides. *Studies in Second Language Acquisition*.
- Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, 35(4), 655-687. <https://doi.org/10.1017/S0272263113000399>
- Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912. <https://doi.org/10.1111/lang.12079>

- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118-137. <https://doi.org/10.1016/j.intell.2015.09.002>
- Shah, J., & Higgins, E. T. (2001). Regulatory concerns and appraisal efficiency: The general impact of promotion and prevention. *Journal of Personality and Social Psychology*, 80(5), 693-705. <https://doi.org/10.1037/0022-3514.80.5.693>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. <https://doi.org/10.1007/s11336-008-9101-0>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *p*-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666-681. <https://doi.org/10.1177/1745691614553988>
- Stark, P., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFROA.v1>
- Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 66-97). Bristol, UK: Multilingual Matters.
- Taylor, F. (2013). *Self and identity in adolescent foreign language learning*. Bristol, UK: Multilingual Matters.
- Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, 39(4), 681-709. <https://doi.org/10.1017/S0272263116000243>
- Thompson, A. S., & Vásquez, C. (2015). Exploring motivational profiles through language learning narratives. *Modern Language Journal*, 99(1), 158-174. <https://doi.org/10.1111/modl.12187>
- Thorsen, C., Henry, A., & Cliffordson, C. (2017). The case of a missing person? The current L2 self and the L2 motivational self system. *International Journal of Bilingual Education and Bilingualism*, Advance online publication.
- Ushioda, E. (2011). Language learning motivation, self and identity: Current theoretical perspectives. *Computer Assisted Language Learning*, 24(3), 199-210. <https://doi.org/10.1080/09588221.2010.538701>
- Wang, F. (2008). Motivation and English achievement: An exploratory and confirmatory factor analysis of a new measurement for Chinese students of English learning. *North American Journal of Psychology*, 10(3), 633-646.
- Wolters, C. A., & Taylor, D. J. (2012). A self-regulated learning perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 635-651). Boston, MA: Springer.

You, C., Dörnyei, Z., & Csizér, K. (2016). Motivation, vision, and gender: A survey of learners of English in China. *Language Learning*, 66(1), 94-123. <https://doi.org/10.1111/lang.12140>

## List of studies included in the meta-analysis

- Al Qahtani, A. F. A. (2015). *Relationships between intercultural contact and L2 motivation for a group of undergraduate Saudi students during their first year in the UK*. Unpublished doctoral dissertation, University of Leeds, UK.
- Al-Hoorie, A. H. (2016). Unconscious motivation. Part II: Implicit attitudes and L2 achievement. *Studies in Second Language Learning and Teaching*, 6, 619-649.
- Alshahrani, A. A. S. (2016). L2 Motivational self system among Arab EFL learners: Saudi perspective. *International Journal of Applied Linguistics and English Literature*, 5, 145-152.
- Al-Shehri, A. S. (2009). Motivation and vision: The relation between the ideal L2 self, imagination and visual style. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 164-171). Bristol, UK: Multilingual Matters.
- Asker, A. (2011). *Future self-guides and language learning engagement of English-major secondary school students in Libya: Understanding the interplay between possible selves and the L2 learning situation*. Unpublished doctoral dissertation, University of Birmingham, UK.
- Busse, V. (2013). An exploration of motivation and self-beliefs of first year students of German. *System*, 41(2), 379-398.
- Calvo, E. T. (2015). *Language learning motivation: The L2 motivational self system and its relationship with learning achievement*. Unpublished MA dissertation, University of Barcelona, Spain.
- Dörnyei, Z., & Chan, L. (2013). Motivation and vision: An analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, 63, 437-462.
- Eid, J. (2008). *Determining the relationship between visual style, imagination, the L2 Motivational self system and the motivation to learn English, French and Italian*. Unpublished MA dissertation, University of Nottingham, UK.
- Ghanizadeh, A., Eishabadi, N., & Rostami, S. (2016). Motivational dimension of willingness to communicate in L2: The impacts of criterion measure, ideal L2 self, family influence, and attitudes to L2 culture. *International Journal of Research Studies in Education*, 5, 13-24.
- Ghanizadeh, A., & Rostami, S. (2015). A Dörnyei-inspired study on second language motivation: A cross-comparison analysis in public and private contexts. *Psychological Studies*, 60, 292-301.
- Huang, H-T., & Chen, I-L. (2016). L2 selves in motivation to learn English as a foreign language: The case of Taiwanese adolescents. In M. T. Apple, D. Da Silva & T. Fellner (Eds.), *L2 selves and motivations in Asian contexts* (pp. 51-69). Bristol, UK: Multilingual Matters.
- Islam, M., Lamb, M., & Chambers, G. (2013). The L2 motivational self system and national interest: A Pakistani perspective. *System*, 41, 231-244.
- Iwaniec, J. (2014). *Motivation to learn English of Polish gymnasium pupils*. Unpublished doctoral dissertation, University of Lancaster, UK.
- Jiang, Y. (2013). *Gender differences and the development of L2 English learners' L2 motivational self system and international posture in China*. Unpublished doctoral dissertation, Birkbeck, University of London, UK.

- Khany, R., & Amiri, M. (2016). Action control, L2 motivational self system, and motivated learning behavior in a foreign language learning context. *European Journal of Psychology of Education*, 1-17.
- Kim, Y.-K., & Kim, T.-Y. (2011). The effect of Korean secondary school students' perceptual learning styles and ideal L2 self on motivated L2 behavior and English proficiency. *Korean Journal of English Language and Linguistics*, 11, 21-42.
- Kim, T.-Y., & Kim, Y.-K. (2014). A structural model for perceptual learning styles, the ideal L2 self, motivated behavior, and English proficiency. *System*, 46, 14-27.
- Lake, J. (2013). Positive L2 self: Linking positive psychology with L2 motivation. In M. T. Apple, D. Da Silva & T. Fellner (Eds.), *Language learning motivation in Japan* (pp. 225-244). Bristol, UK: Multilingual Matters.
- Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, 62, 997-1023.
- Lasagabaster, D. (2016). The relationship between motivation, gender, L1 and possible selves in English-medium instruction. *International Journal of Multilingualism*, 13, 315-332.
- Moskovsky, C., Assulaimani, T., Racheva, S., & Harkins, J. (2016). The L2 motivational self system and L2 achievement: A study of Saudi EFL learners. *Modern Language Journal*, 100, 641-654.
- Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System*, 38(3), 467-479.
- Papi, M., & Abdollahzadeh, E. (2012). Teacher motivational practice, student motivation, and possible L2 selves: An examination in the Iranian EFL context. *Language Learning*, 62, 571-594.
- Polat, N. (2014). The interaction of the L2 Motivational self system with socialisation and identification patterns and L2 accent attainment. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 268-285). Bristol, UK: Multilingual Matters.
- Ryan, S. (2009). Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 120-143). Bristol, UK: Multilingual Matters.
- Shahbaz M., & Liu, Y. (2012). Complexity of L2 motivation in an Asian ESL setting. *Porta Linguarum*, 18, 115-131.
- Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 66-97). Bristol, UK: Multilingual Matters.
- Tan, T. G., Lim, T. H., & Hoe, F. T. (2017). Analyzing the relationship between L2 motivational self system and achievement in Mandarin. *International Academic Research Journal of Social Science*, 3(1), 104-108.
- Yun, S., Hiver, P., & Al-Hoorie, A. H. (in press). Academic buoyancy: Construct validation and a test of structural relations. *Studies in Second Language Acquisition*.

## APPENDIX B

## Characteristics of studies included in the meta-analysis

Study	Sample size	Target language	Learner level	Learner gender	Context	Criterion measure
Al Qahtani (2015)	257	English	University	Mixed	M. East	Intended effort
Al-Hoorie (2016)	311	English	University	Male	M. East	Mixed
Alshahrani (2016)	397	English	University	Male	M. East	Intended effort
Al-Shehri (2009)	200	English	Mixed	Mixed	M. East	Intended effort
Asker (2011)	126	English	Secondary	Mixed	M. East	Intended effort
Busse (2013)	94	German	University	Mixed	Europe	Intended effort
Calvo (2015)	29	English	Secondary	Mixed	Europe	Grades
Dörnyei & Chan (2013)	172	Mixed	Secondary	Mixed	Asia	Mixed
Eid (2011)	93	Mixed	Secondary	Mixed	Europe	Mixed
Ghanizadeh et al. (2016)	160	English	University	Mixed	M. East	Intended effort
Ghanizadeh & Rostami (2015)	905	English	Mixed	Mixed	M. East	Intended effort
Huang & Chen (2016)	1,698	English	Secondary	Mixed	Asia	Intended effort
Islam et al. (2013)	975	English	University	Mixed	M. East	Intended effort
Iwaniec (2014)	236	English	Secondary	Mixed	Europe	Intended effort
Jiang (2013)	240	English	University	Mixed	Asia	Intended effort
Khani & Amiri (2016)	510	English	Secondary	Mixed	M. East	Intended effort
Kim & Kim (2011)	495	English	Secondary	Mixed	Asia	Intended effort
Kim & Kim (2014)	2,239	English	Secondary	Mixed	Asia	Intended effort
Lake (2013)	224	English	University	Female	Asia	Mixed
Lamb (2012)	527	English	Secondary	Mixed	Asia	Mixed
Lasagabster (2016)	189	English	University	Mixed	Europe	Intended effort
Moskovsky et al. (2016)	360	English	University	Mixed	M. East	Mixed
Papi (2010)	1,011	English	Secondary	Mixed	M. East	Mixed
Papi & Abdollahzadeh (2012)	460	English	Secondary	Male	M. East	Behavior
Polat (2014)	88	English	Secondary	Mixed	M. East	Pronunciation
Ryan (2009)	2,397	English	Mixed	Mixed	Asia	Intended effort
Shahbaz & Liu (2012)	547	English	University	Mixed	M. East	Intended effort
Shahbaz & Yongbing (2015)	547	English	University	Mixed	M. East	Intended effort
Taguchi et al. (2009)	4,891	English	Mixed	Mixed	Mixed	Intended effort
Tan et al. (2017)	142	English	University	Mixed	Asia	Grades
You & Dörnyei (2016)	10,413	English	Mixed	Mixed	Asia	Intended effort
Yun et al. (in press)	787	English	University	Mixed	Asia	Grades



APPENDIX C

List of studies excluded for incomplete reporting of Pearson correlation results

Studies reporting regression coefficients:

- Csizér, K., & Lukács, G. (2010). The comparative analysis of motivation, attitudes and selves: The case of English and German in Hungary. *System*, 38(1), 1-13.
- Khan, M. R. (2015). Analyzing the relationship between L2 motivational selves and L2 achievement: A Saudi perspective. *International Journal of English Language Teaching*, 2(1), 68-75.
- Kim, T.-Y., & Kim, Y.-K. (2014). EFL students' L2 motivational self system and self-regulation: Focusing on elementary and junior high school students in Korea. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 87-107). Bristol, UK: Multilingual Matters.
- Kormos, J., & Csizér, K. (2008). Age-related differences in the motivation of learning English as a foreign language: Attitudes, selves, and motivated learning behavior. *Language Learning*, 58(2), 327-355.
- Kormos, J., & Csizér, K. (2010). A comparison of the foreign language learning motivation of Hungarian dyslexic and non-dyslexic students. *International Journal of Applied Linguistics*, 20(2), 232-250.
- Li, Q. (2014). Differences in the motivation of Chinese learners of English in a foreign and second language context. *System*, 42, 451-461.
- Mezei, G. (2014). The effect of motivational strategies on self-related aspects of student motivation and second language learning. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 289-309). Bristol, UK: Multilingual Matters.
- Nagle, C. (2014). *A longitudinal study on the role of lexical stress and motivation in the perception and production of L2 Spanish stop consonants*. Unpublished doctoral dissertation, Georgetown University, USA.
- Papi, M., & Teimouri, Y. (2013). Dynamics of selves and motivation: A cross-sectional study in the EFL context of Iran. *International Journal of Applied Linguistics*, 22(3), 287-309.
- Polat, N., & Schallert, D. L. (2013). Kurdish adolescents acquiring Turkish: Their self-determined motivation and identification with L1 and L2 communities as predictors of L2 accent attainment. *The Modern Language Journal*, 97(3), 745-763.
- Sugita McEown, M., Noels, K. A., & Chaffee, K. E. (2014). At the interface of the Socio-Educational Model, Self-Determination Theory and the L2 Motivational self system models. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 19-50). Bristol, UK: Multilingual Matters.
- Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, 39(4), 681-709.
- Xie, Y. (2014). L2 self of beginning-level heritage and nonheritage postsecondary learners of Chinese. *Foreign Language Annals*, 47(1), 189-203.

Studies reporting SEM path coefficients:

- Aubrey, S. (2014). Development of the L2 Motivational self system: English at a university in Japan. *JALT Journal*, 36(2), 153-174.

- Aubrey, S., & Nowlan, G. P. (2013). Effect of intercultural contact on L2 motivation: A comparative study. In M. T. Apple, D. Da Silva & T. Fellner (Eds.), *L2 selves and motivations in Asian contexts* (pp. 129-151). Bristol, UK: Multilingual Matters.
- Csizér, K., & Kormos, J. (2009). Learning experiences, selves and motivated learning behaviour: A comparative analysis of structural models for Hungarian secondary and university learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 98-119). Bristol, UK: Multilingual Matters.
- Gu, M., & Cheung, D. S. (2016). Ideal L2 self, acculturation, and Chinese language learning among South Asian students in Hong Kong: A structural equation modelling analysis. *System*, 57, 14-24.
- Kormos, J., & Csizér, K. (2014). The interaction of motivation, self-regulatory strategies, and autonomous learning behavior in different learner groups. *TESOL Quarterly*, 48, 275-299.
- Kormos, J., Kiddle, T., & Csizér, K. (2011). Goals, attitudes and self-related beliefs in second language learning motivation: An interactive model of language learning motivation. *Applied Linguistics*, 32(5), 495-516.
- Magid, M. (2011). *A validation and application of the L2 Motivational self system among Chinese learners of English*. Unpublished doctoral dissertation, University of Nottingham, UK.
- Munezane, Y. (2016). Motivation, ideal self and willingness to communicate as the predictors of observed L2 use in the classroom. *EUROSLA Yearbook*, 16, 85-115.
- Taguchi, T. (2013). Motivation, attitudes and selves in the Japanese context: A mixed methods approach. In M. T. Apple, D. Da Silva & T. Fellner (Eds.), *L2 selves and motivations in Asian contexts* (pp. 169-188). Bristol, UK: Multilingual Matters.
- Ueki, M., & Takeuchi, O. (2013). Forming a clearer image of the ideal L2 self: The L2 Motivational self system and learner autonomy in a Japanese EFL context. *Innovation in Language Learning and Teaching*, 7(3), 238-252.
- Visgatis, B. L. (2014) *English-related out-of-class time use by Japanese university students*. Unpublished doctoral dissertation, Temple University, USA.
- Studies reporting chi-square and Kendall's tau, respectively:
- Georgiadou, E. S. (2016). The role of proficiency, speaking habits and error-tolerance in the self-repair behaviour of Emirati EFL learners. *The Asian EFL Journal Quarterly*, 18(4), 104-126.
- Kim, T.-Y. (2009). Korean elementary school students' perceptual learning style, ideal L2 self, and motivated behavior. *Korean Journal of English Language and Linguistics*, 9(3), 461-486.