Research Methods in Applied Linguistics 5 (2026) 100332

Contents lists available at ScienceDirect

Research Methods in Applied Linguistics

journal homepage: www.elsevier.com/locate/rmal

Re-assessing validity evidence for the L2 motivational self system: a response to McClelland and Larson-hall (2025)

Mostafa Papi [a,*], Wenting Song [a], Hadya Soliman [b]

[a] *Florida State University, USA*
[b] *Suez Canal University, Egypt*

ARTICLE INFO

Keywords:
Methodological reform
L2MSS
Ideal L2 Self
Discriminant validity
Predictive validity
Construct validity

ABSTRACT

This study addressed concerns raised by McClelland and Larson-Hall (2025) about the construct validity of the L2 Motivational Self System (L2MSS) and, by extension, the validity of numerous studies that have used this model and the L2MSS questionnaire developed by Taguchi et al. (2009). Using the original data from Taguchi et al. (2009), this study examined the construct, convergent, discriminant, and predictive validity of the L2MSS constructs using various statistical techniques. Confirmatory factor analyses, average variance extracted, Heterotrait-Monotrait Ratio, Pearson correlations, multiple regression, and reliability analyses confirmed the construct, discriminant, convergent, and predictive validity of the scales. Additionally, both the Ideal L2 Self and Integrativeness significantly predicted Intended Effort, but the effect of the Ideal L2 Self was significantly stronger than that of Integrativeness, supporting one of the core claims of the L2MSS tradition. Taken together, these findings suggest that the L2MSS remains a valid and useful framework, and that calls for abandoning it are not supported by the present evidence. We therefore encourage the continued use of the L2MSS, particularly through theoretically refined frameworks such as the 2×2 Model of Future Self-Guides, alongside ongoing theoretical and methodological development.

In a critical commentary, McClelland and Larson-Hall (2025) argued that the L2 Motivational Self System (L2MSS; Dörnyei, 2009) and especially its key construct, Ideal L2 Self, should no longer be used in research on L2 motivation. They revisited some foundational studies and contended that the methodological practices used have undermined empirical support for the model. Below are two of the main arguments they made in support of their claims:

1) The methods for establishing construct validity are flawed. As early L2MSS studies often "consistently and systematically omitted" (p. 1122) exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Furthermore, when SEM or CFA was used, it was not conducted properly. More specifically, researchers failed to compare competing models, employed excessive post-hoc modifications, and set high thresholds for discriminant validity. McClelland and Larson-Hall (2025) argued that "over nearly 20 years, researchers in the tradition simply assumed the validity of their instruments, without subjecting them to the rigorous testing usually expected of scientific inquiry" (p. 1123).

* Corresponding author at: Florida State University, 1114 W. Call St, G129, Stone Building, School of Teacher Education, Tallahassee, FL 32306, USA.
E-mail address: mpapi@fsu.edu (M. Papi).

https://doi.org/10.1016/j.rmal.2026.100332
Received 27 March 2026; Received in revised form 5 June 2026; Accepted 6 June 2026
Available online 11 June 2026

2) The foundational L2MSS studies failed to demonstrate that the Ideal L2 Self was a more powerful motivational variable than Integrativeness, as they relied on correlations (Ryan, 2009) or excluded Integrativeness from the models (Taguchi et al., 2009).

McClelland and Larson-Hall (2025) highlighted the importance of using factor-analytic methods to examine construct, discriminant, and convergent validity. However, their characterization of foundational L2MSS studies, such as Taguchi et al. (2009), is not entirely accurate. This is because seminal studies, such as those by Taguchi et al. (2009) and Papi (2010), did employ CFA, although they did not report the results due to space limitations and the conventions of the time. In addition, the modifications were primarily made at the structural stage, which was expected due to the sophisticated nature of the models being tested. McClelland and Larson-Hall (2025) are, however, right to raise questions about the validity of the L2MSS studies if such important information was left out of published reports. Therefore, this study aims to address those concerns by re-analyzing data from Taguchi et al. (2009) to test the construct, convergent, discriminant, and predictive validity of the L2MSS components.

Based on a series of exploratory studies in Hungary (e.g., Dörnyei, 2006), Taguchi et al. (2009) developed and validated a questionnaire to examine the components of the L2MSS, using data collected from nearly 5000 participants in Japan, China, and Iran. Although there were minor context-based differences in the questionnaires used across the three contexts, the questionnaire scales included largely the same items, and the differences were also conceptually aligned with the construct definitions (see Taguchi et al., 2009). McClelland and Larson-Hall (2025) further argued that Taguchi et al. (2009) did not follow the best-practice model-comparison guidelines proposed by Schoonen (2015). Specifically, instead of testing competing theoretical models within a single dataset, Taguchi et al. (2009) tested one model across three separate datasets. Because access was limited to the Iranian dataset, this response will focus on re-analyzing those data while addressing the model-comparison concerns raised by McClelland and Larson-Hall (2025).

Taguchi et al. (2009) collected data from 2029 Iranian learners of English as a foreign language (EFL; 1137 females, 892 males) including middle school ($N = 1309$) and university students (English major = 394, non-English majors = 325) from different provinces all across Iran (see Taguchi et al., 2009, for more detailed descriptions) to validate the original questionnaire that has been frequently used in the L2MSS studies and was the focus of McClelland and Larson-Hall's (2025) critique.

1. This response

In the re-analysis of the data to address the first concern regarding the method for establishing construct validity, we carried out three steps using factor analysis. First, we conducted a CFA on the three components of the L2MSS model and reported all necessary standard indices. Second, we compared different CFA models to evaluate discriminant and convergent validity. Third, we reported models without any post-hoc modifications, addressing concerns about potential overuse of model modifications. To examine the discriminant validity of the L2MSS constructs in relation to Intended Effort, we conducted Pearson correlations and CFA analyses. Finally, to address the claim regarding the lack of evidence showing the superiority of the Ideal L2 Self over Integrativeness, we ran a model-comparison CFA and multiple regression analyses to compare their predictive validity. It must be noted that we do not aim to argue for the continued use of the L2MSS in its original form; rather, we recommend its revised version (e.g., Papi et al., 2019; Papi & Khajavy, 2021). However, the purpose of the present response is to evaluate McClelland and Larson-Hall's (2025) claim against the validity of the original L2MSS, and to examine whether their critique warrants dismissing the large body of research that has employed the model.

2. The instrument

We used data collected by Taguchi et al. (2009) in Iran. The original questionnaire included items measuring several factors, including but not limited to Ideal L2 Self, Ought-to L2 Self, Attitudes toward L2 Learning Experience (L2LE), Integrativeness, and Intended Effort (criterion measure) (Appendix A), as well as demographic questions. The items were presented in both question and statement forms, and all factors were measured on a Likert scale from 1 to 6. For question-format items, 1 indicated '*not at all'* and 6 indicated '*very much,'* while for statement-format items, 1 corresponds to '*strongly disagree'* and 6 corresponds to '*strongly agree'*. A total of 76 items were administered to participants.

More specifically, for the L2MSS framework, each component was assessed with six items. Both Ideal L2 Self and Ought-to L2 Self were measured through statement-type items. An example item for the Ideal L2 Self is "*I can imagine myself speaking English as if I were a native speaker of English*," and an example for Ought-to L2 Self is "*I study English because close friends of mine think it is important*." Unlike the previous two scales, the items of the L2LE scale were framed as questions, for instance, "*Do you always look forward to English classes*?" Similarly, Integrativeness was measured using question-format items (e.g., *How much would you like to become similar to the people who speak English?*). Finally, the Intended Effort was measured through six items in a statement-type format (e.g., *I would like to spend lots of time studying English*).

3. Argument 1: construct, discriminant, and convergent validity of L2MSS

Using the *Lavaan* package in the R program (Rosseel, 2012), we conducted a series of CFAs to validate the factorial structure of the L2MSS model. The main model specified three latent variables: Model 1: Ideal L2 Self, Ought-to L2 Self, and L2LE. To check the discriminant validity of the constructs, we compared the model with alternative models. These models included three two-factor models: Model 2: Ideal L2 Self plus one factor with Ought-to Self and L2LE merged; Model 3: Ought-to Self plus one factor with Ideal L2 Self and L2LE merged; and Model 4: L2LE plus one factor with Ideal L2 Self and Ought-to L2 Self merged. In addition to the

Table 1
Confirmatory factor results for L2MSS and alternative models.

	$\chi^2(df)$	Standardized Loading Range	CFI	TLI	AIC	BIC	RMSEA [90%CI]	SRMR	$\Delta\chi^2$ *(df)*
L2MSS	623.00*** (132)	[.402, 0.794]	.952	.944	123,403.86	123,723.93	.043[.039, 0.046]	.036	
Model 2	2880.413 (134)	[.079, 0.781]	. 729	.691	125,657.27	125,966.12	.101[.097, 0.104]	.099	2257.4***(2)
Model 3	1377.51 (134)	[.405, 0.747]	.877	.860	124,154.37	124,463.21	.068[.064, 0.071]	.053	754.51***(2)
Model 4	2612.56 (134)	[.276, 0.793]	.756	.721	125,389.42	125,698.27	.095[.092, 0.099]	.090	1989.6***(2)

Note. Model 2: Ideal L2 Self & (Ought-to Self + L2LE); Model 3: Ought-to Self & (Ideal L2 Self + L2LE); and Model 4: L2LE & (Ideal L2 Self + Ought-to L2 Self). CFI = Comparative Fit Index, TLI = Tucker–Lewis Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, $\Delta\chi^2$ *(df)* = Chi-square difference test.
***$p < .001$.

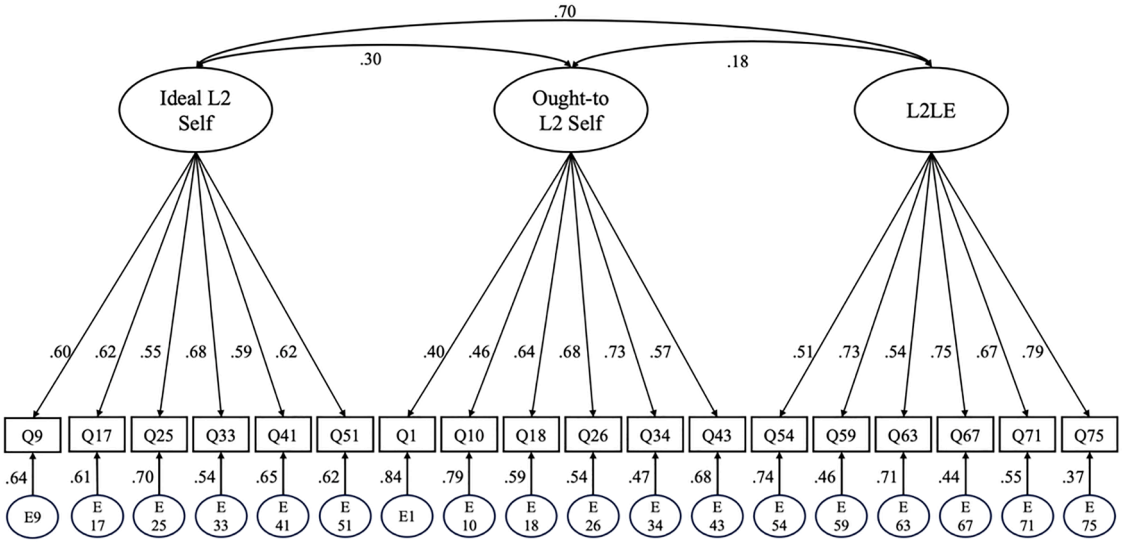

Fig. 1. Confirmatory factor analysis results for the complete L2MSS model.

aforementioned models, we also used a separate model to test the discriminant validity of the Integrativeness and Ideal L2 Self. We examined the model fit using multiple indices: The chi-square test, Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). When CFI and TLI values are $< .90$, the model fit is poor; A value between 0.90 and 0.94 indicates adequate fit; and a value of > 0.95 indicates excellent fit. RMSEA and SRMR < 0.10 is considered poor, 0.08 – 0.10 is mediocre, 0.06 – 0.08 is acceptable, and $\leq .05$ shows excellent fit (Brown, 2015; Kline, 2023). Additionally, we conducted a chi-square difference test to evaluate whether the alternative models provide a significantly better fit than the primary model.

Table 1 presents the CFA results for the L2MSS and its chi-square difference test relative to alternative models. For the complete L2MSS model (see Fig. 1), the chi-square test result was significant, $\chi^2(132) = 623.00$, $p < .001$, which is not uncommon in large samples. However, the remaining indices supported good model fit, CFI = 0.952, TLI = 0.944, RMSEA = 0.043, 90% CI [.039, 0.046], SRMR = 0.036. By contrast, when testing the alternative models (Model 2 & 3), all the indices showed poorer fit. More specifically, Model 3 (Ought-to L2 Self & Ideal L2 Self + L2LE) yielded CFI and TLI below 0.90, indicating poor fit, while RMSEA and SRMR were acceptable or good. Additionally, despite the significant chi-square results, Models 2 and 4 had CFI and TLI values below 0.90. Furthermore, the RMSEA values of the other two models were above 0.08, with even the lower bound of the 90% confidence intervals above 0.08. The SRMR values were also above 0.08, suggesting mediocre fit for the data. Finally, in addition to the poor fit indices of the alternative models, the series of chi-square difference tests were all significant, and the AIC and BIC values for the alternative models were higher than those for the complete model. Taken together, these results support the superiority of the full L2MSS model without the need for any post-hoc modifications. They also address the concerns raised by McClelland and Larson-Hall (2025) regarding the overuse of post-hoc model modifications, as the L2MSS model demonstrated good fit without any statistically driven adjustments.

Next, rather than excluding Integrativeness from the model, we directly tested the discriminant validity alongside the Ideal L2 Self using factor analysis. Table 2 presents the CFA results for the bifactorial model of Integrativeness and Ideal L2 Self, compared with a unifactorial model with both scales merged (i.e., loading all items onto a single variable). Overall, the CFA results for the bifactorial model showed an acceptable to good fit. Although the chi-square test result was significant, $\chi^2(26) = 217.00$, $p < .001$, other fit indices supported model fit, CFI = 0.952, TLI = 0.933, RMSEA = 0.06 with 90% CI [.053, 0.068], SRMR = 0.032. For the unifactorial model, in

Table 2
Confirmatory factor results for Ideal L2 Self and Integrativeness models.

Model	$\chi^2(df)$	Standardized Loading Range	CFI	TLI	AIC	BIC	RMSEA [90%CI]	SRMR	$\Delta\chi^2$ (df)
Bifactorial	217.00 *** (26)	[.41, 0.71]	.95	.93	61,003.89	61,161.12	.060 [.053, 0.068]	0.032	
Unifactorial	333.03 *** (27)	[.36, 0.66]	.92	.89	61,118.06	61,269.67	.075 [.068, 0.082]	0.040	116.16***(1)

Note. CFI = Comparative Fit Index, TLI = Tucker–Lewis Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, $\Delta\chi^2$ *(df)* = Chi-square difference test.
***$p < .001$.

Table 3
Reliability analysis results.

	α	ω	CR
Ideal L2 Self	.78	.84	.78
Ought-to L2 Self	.75	.84	.76
L2 Learning Experience	.82	.88	.83

Note. α = Cronbach Alpha, ω = McDonald's Omega, CR = Composite Reliability.

addition to the significant chi-square result, the TLI value was below 0.90, the RMSEA and SRMR values were higher, the values of AIC and BIC were higher than the bifactor model, and the chi-square difference test was significant, confirming that the bifactorial model was a better model than the unifactorial model. These results support the discriminant validity of the Ideal L2 Self and the Integrativeness, which were only moderately correlated ($r = 0.53$).

The reliability indices were reported in Table 3. More specifically, the Cronbach's alpha values, ranging from 0.75 to 0.82, indicated acceptable to good internal consistency for all three subscales of the L2MSS. In response to the concern regarding the exclusive use of Cronbach's Alpha, we computed McDonald's Omega, which showed good reliability, ranging from 0.84 to 0.88. Furthermore, the composite reliability values were all above the recommended threshold of 0.70, further supporting the internal consistency and convergent validity of the scales.

To further address the discriminant validity of the three core components of L2MSS, we computed Average Variance Extracted (AVE) (see Table 4) and the Heterotrait-Monotrait Ratio of Correlations (HTMT). Fornell and Larcker's (1981) criterion suggested that the AVE for a latent variable should be larger than its squared correlation with other variables to support discriminant validity. The AVE for Ought-to L2 Self (0.35) was larger than its squared correlation with Ideal L2 Self ($R^2 = 0.09$) and L2LE (R^2=0.03), showing good discriminant validity. For the Ideal L2 Self, the AVE (0.37) was larger than its squared correlation with Ought-to Self (R^2= 0.09), but smaller than L2LE (R^2=0.49), indicating a potential overlap between Ideal L2 Self and L2LE. However, Henseler et al. (2015) commented on the insufficiency of using Fornell and Larcker's criterion to examine discriminant validity; instead, they suggested the HTMT, which is the ratio of the average of between-construct correlations (i.e., heterotrait-heteromethod correlations) to within-construct correlations (i.e., monotrait-heteromethod correlations). The HTMT values were below the 0.85 threshold, indicating an adequate distinction among the constructs. Specifically, HTMT values were 0.29 (Ideal–Ought), 0.70 (Ideal–L2LE), and 0.18 (Ought–L2LE; Voorhees et al., 2016).

Table 4
Average variance extracted and interconstruct correlations.

	Ideal L2 Self	Ought-to L2 Self	L2 Learning Experience	AVE
Ideal L2 Self	-	-	-	.37
Ought-to L2 Self	.30	-	-	.35
L2 Learning Experience	.70	.18	-	.45

Note. AVE = Average Variance Extracted.

Table 5
Descriptive statistics for the target variables.

	Ideal L2 Self	Ought-to L2 Self	L2 Learning Experience	Integrativeness	Intended Effort
M	4.45	3.39	4.34	4.63	4.54
Median	4.5	3.5	4.5	4.67	4.67
SD	1.02	1.07	1.11	1.05	1.00
Minimum	1.00	1.00	1.00	1.00	1.00
Maximum	6.00	6.00	6.00	6.00	6.00

4. Argument 2: predictive validity problem

McClelland and Larson-Hall (2025) argued that the statistical relationships between L2MSS components and criterion measures could be "largely an artifact of overlapping wording in the measurement and criterion scales" (p. 1127), suggesting a lack of discriminant validity. To address this concern, we conducted Pearson correlation, and CFA analyses. The descriptives of the five latent variables (i.e., Ideal L2 Self, Ought-to L2 Self, L2LE, Integrativeness, and Intended Effort) are presented in Table 5. Pearson correlations among the L2MSS components (Ideal L2 Self, Ought-to L2 Self, L2LE), Integrativeness, and Intended Effort are presented in Table 6. The strongest significant positive correlation was found between L2LE and Intended Effort ($r = 0.71$, $p < .001$), which is significantly lower than the 0.84 reported by Al-Hoorie et al. (2024). According to Brown (2015), "in applied research, a factor correlation that equals or exceeds 0.85 is often used as the cutoff criterion for problematic discriminant validity" (p. 147; see also Rönkkö & Cho, 2022). However, McClelland and Larson-Hall's argument appears to conflate effect-size benchmarks with discriminant-validity thresholds when extending Al-Hoorie et al.'s (2025) argument by suggesting that correlations of .60 or even .40 represent "conventional benchmarks" for identifying discriminant validity concerns. On this basis, they contended that a cutoff point of 0.85 was an "unusually high threshold […] for identifying problems of discriminant validity between scales" (p. 1127). To further support the argument that a correlation of 0.71 is not indicative of a discriminant validity problem, we ran another model-comparison CFA with L2LE and Intended Effort as separate and merged constructs. The results showed that the two-factor model exhibited a better fit (CFI = 0.962; TLI = 0.944; RMSEA = 0.055, 90% CI [.050, 0.060]) than the merged model (CFI = 0.939; TLI = 0.912; RMSEA = 0.069, 90% CI [.064, 0.074]). Model comparison using AIC also favored the two-factor structure (AIC = 452.73) over the one-factor structure (AIC = 648.94). The two-factor model provided a substantially better fit and reflected a more accurate representation of the latent structure than the merged model, supporting the discriminant validity of the constructs.

Among other correlations, Ideal L2 Self and L2LE ($r = 0.56$, $p < .001$) were strongly correlated, suggesting that the higher the L2 learner's Ideal L2 Self scores, the more positive their L2 Learning Experience tends to be. On the other hand, the analysis revealed a weak correlation between Ought-to L2 Self and L2LE ($r = 0.14$, $p < .001$). This finding suggests that social pressures, such as disappointing parents or letting significant others down, are weakly associated with positive attitudes toward L2 learning. Finally, Ideal L2 Self ($r = 0.61$) and Integrativeness ($r = 0.58$) both showed strong correlations with Intended Effort. Considering the high correlation between Ideal L2 Self and Integrativeness ($r = 0.53$), multiple regression analyses were run to explore the unique contribution of each variable, as well as those of the other components of the L2MSS, to Intended Effort.

5. Argument 3: ideal L2 self vs. Integrativeness: which one wins?

Table 7 presents the regression analysis results. The model, which included Ideal L2 Self and Integrativeness as predictors, was statistically significant, $F(2, 2026) = 877.13$, $p < .001$, explaining 46.4% of the variance in Intended Effort ($R^2 = 0.464$, adjusted $R^2 = 0.464$). Both Integrativeness ($\beta = 0.362$, $t = 18.91$, $p < .001$) and Ideal L2 Self ($\beta = 0.417$, $t = 21.80$, $p < .001$) were significant positive predictors of Intended Effort. Ideal L2 Self emerged as the stronger predictor, as indicated by its larger standardized regression coefficient. Collinearity diagnostics suggested no multicollinearity concerns (VIF = 1.38 for both predictors).

To compare the relative predictive strength of Ideal L2 Self and Integrativeness, their unstandardized regression coefficients were compared using a custom contrast (LMATRIX). The results of the comparison test showed that Ideal L2 Self was a statistically stronger predictor of Intended Effort than Integrativeness, $p = 0.046$, 95% CI [.001, 0.128], supporting Dörnyei's (2009) claim for the stronger motivational power of the Ideal L2 Self.

Finally, we conducted a multiple regression analysis to predict Intended Effort from the three L2MSS components (Ideal L2 Self, Ought L2 Self, L2LE) and Integrativeness. Table 8 presents the regression analysis results. The model, which included all four predictors, was statistically significant, $F(4, 2024) = 708.29$, $p < .001$, explaining 58.3% of the variance in Intended Effort ($R^2 = 0.583$, adjusted $R^2 = 0.582$). All predictors were statistically significant, with L2LE being the strongest predictor of Intended Effort, ($\beta = 0.479$, $B = 0.432$, $SE = 0.018$, $t = 23.72$, $p < .001$), followed by Ideal L2 Self ($\beta = 0.271$, $B = 0.266$, $SE = 0.018$, $t = 15.06$, $p < .001$), Integrativeness ($\beta = 0.112$, $B = 0.106$, $SE = 0.019$, $t = 5.64$, $p < .001$), and Ought-to L2 Self ($\beta = 0.040$, $B = 0.055$, $SE = 0.014$, $t = 4.00$, $p < .001$). Collinearity diagnostics indicated no multicollinearity concerns, with VIF values ranging from 1.05 to 1.94. These results indicate that the Ideal L2 Self and L2LE make unique contributions to the outcome variable, thereby supporting the discriminant and predictive validity of these constructs.

Furthermore, to compare the relative predictive strength of the Ideal L2 Self and Integrativeness in this model, we conducted another custom contrast (LMATRIX) on their unstandardized regression coefficients. Similar to the previous model, the results showed

Table 6
Pearson correlation coefficients for L2MSS.

	Ideal L2 Self	Ought-to L2 Self	L2 Learning Experience	Integrativeness	Intended Effort
Ideal L2 Self	-				
Ought-to L2 Self	.22***	-			
L2 Learning Experience	.56***	.14***	-		
Integrativeness	.53***	.17***	.66***	-	
Intended Effort	.61***	.21***	.71***	.58***	-

Note. *** $p < .001$.

Table 7
Multiple regression analysis with Integrativeness and Ideal L2 Self as predictors.

Predictor	B	SE	β	*t*	*p*	95% CI for B	VIF
Intercept	1.129	0.083	—	13.57	< 0.001	[0.966, 1.292]	—
Integrativeness	0.344	0.018	.362	18.91	< 0.001	[0.308, 0.380]	1.38
Ideal L2 Self	0.409	0.019	.417	21.80	< 0.001	[0.372, 0.445]	1.38

Table 8
Multiple regression analysis with L2MSS components and Integrativeness as predictors.

Predictor	B	SE	*β*	*t*	*p*	95% CI for B	VIF
Constant	0.799	0.080	—	10.12	< 0.001	[0.643, 0.956]	—
Ideal L2 Self	0.266	0.018	.271	15.06	< 0.001	[0.231, 0.300]	1.576
Ought-to L2 Self	0.055	0.014	.040	4.00	< 0.001	[0.029, 0.083]	1.047
L2 Learning Experience	0.432	0.018	.479	23.72	< 0.001	[0.396, 0.468]	1.937
Integrativeness	0.106	0.019	.112	5.64	< 0.001	[0.069, 0.143]	1.910

that the Ideal L2 Self was a statistically stronger predictor of Intended Effort than Integrativeness ($p < 0.001$, 95% CI [.104, 0.230]). This finding also supports the central claim of the L2MSS that learners' future self-images constitute a more powerful motivational force than the traditional construct of Integrativeness.

6. Conclusion

To address the concerns raised by McClelland and Larson-Hall (2025), we reanalyzed the original questionnaire data collected by Taguchi et al. (2009) from 2029 EFL learners in Iran. The results of the CFA supported the three-factor structure of the L2MSS, indicating that the model fit better than alternative models with merged factors, without requiring any post-hoc modifications to the models representing the measured constructs. In addition, multiple regression analyses showed that all three components of the L2MSS significantly contributed to Intended Effort, supporting their predictive and discriminant validity, as they made unique contributions to the outcome variable without causing multicollinearity (Lawson & Robins, 2021). Taken together, the CFA and multiple regression analyses provide evidence against McClelland and Larson-Hall's (2025) claims that researchers in the L2MSS tradition "simply assumed the validity of their instruments, without subjecting them to the rigorous testing usually expected of scientific inquiry" (p. 1123) and that "Taguchi et al. (2009) missed the opportunity to rigorously test the L2MSS" (p. 1127). Notably, the CFA analyses reported in the present study were also conducted by Taguchi et al. (2009), although they were not reported in the original publication.

Regarding one of the core claims of the L2MSS, that is the superiority of the Ideal L2 Self over Integrativeness, the authors argued that the "conclusion 'that [Ideal L2 Self] might rightfully take over the place of Integrativeness'… is unsupported and speculative" (McClelland & Larson-Hall, 2025, p. 1125). To address this issue more clearly, a CFA with the two variables in a two-factor model showed better fit than a model with the two constructs merged, providing support for their discriminant validity ($r = 0.53$). More importantly, Ideal L2 Self was a significantly stronger predictor of Intended Effort than Integrativeness, as demonstrated by a custom contrast (LMATRIX). These results support one of the core claims of the L2MSS tradition that Ideal L2 Self is a stronger motivational construct than Integrativeness, which appears to be context-dependent and particularly relevant in situations such as immigration or close identification with an L2 community.

One limitation of our study is that the analysis was restricted to the Iranian context, rather than the full dataset used by Taguchi et al. (2009), which also included data from China and Japan. Therefore, the generalizability of the model across cultures might be limited. Nevertheless, our findings provide evidence for the construct and convergent validity of the L2MSS model, while indicating substantial shared variance between L2LE and Intended Effort, although CFA results support their treatment as distinct constructs. These findings provide additional support for the validity of numerous previous studies that have used the questionnaire developed by Taguchi et al. (2009). In addition, these results further suggest the criticisms advanced by McClelland and Larson-Hall (2025) and Al-Hoorie et al. (2024) against L2MSS may be overstated.

The authors also suggested replacing L2MSS with Self-Determination Theory (SDT) (see also Al-Hoorie et al., 2025). However, this characterization overlooks important theoretical distinctions between the two motivational perspectives, which originate from different theoretical traditions and are designed to capture distinct dimensions of human motivation. The L2MSS perspective provides a view of motivation rooted in the classic pleasure-versus-pain tradition in the science of motivation, which diverges from the autonomy- and competence-based tradition presented in theories such as SDT (Deci & Ryan, 1985). These two perspectives address different dimensions of motivation: L2MSS emphasizes the value dimension, and SDT reflects the control dimension (Higgins, 2011; Papi & Hiver, 2020). This means that whereas future selves represent end-states that motivate approaching positive outcomes or avoiding negative outcomes, SDT concerns the degree to which different motives are self-determined.

More specifically, from a regulatory focus perspective (Higgins, 1997), which builds on the self-discrepancy theory (Higgins, 1987), the same goal can be represented in either promotion-focused terms (e.g., gains and non-gains) or prevention-focused terms (e.g., losses and non-losses), and such a division can be harnessed to create qualitative differences in learners' task emotions, engagement, and performance (e.g., Cho, 2021, 2026; Papi, 2018; Papi et al., 2025). Individuals with a promotion focus tend to experience greater

enjoyment and engagement when pursuing goals through eager strategies that emphasize advancement and gains, whereas individuals with a prevention focus tend to experience greater enjoyment and engagement when pursuing goals through vigilant strategies that emphasize responsibility and the avoidance of losses (Freitas & Higgins, 2002; Papi et al., 2025). Thus, from a regulatory focus perspective, enjoyment is not inherently tied to either promotion or prevention focus; rather, it emerges from the fit between individuals' regulatory orientation and the strategic manner in which they pursue their goals. Such qualitative distinctions are addressed more directly within the regulatory focus and future-self perspectives than within SDT, which focuses on how psychological need satisfaction shapes the internalization of motives and their degree of self-determination. In summary, SDT and the L2MSS address complementary dimensions of motivation and should therefore be viewed as theoretically compatible rather than interchangeable. Together, they offer a more comprehensive understanding of language learning motivation than either perspective alone.

The results of this study suggest that the concerns about the construct validity of the L2MSS, while not unfounded, have been overstated by McClelland and Larson-Hall (2025). Accordingly, we encourage researchers to continue employing the L2MSS, either in its original form (Dörnyei, 2009) or, preferably, in its revised form, namely the 2 × 2 Model of Future Self-Guides (e.g., Papi et al., 2019), which offers a more theoretically nuanced representation of L2 future self-guides and is supported by more refined measurement instruments (Papi & Khajavy, 2021). Grounded in Self-Discrepancy Theory (Higgins, 1987), the revised model bifurcates future self-guides by standpoint (own vs. other) and regulatory focus (promotion vs. prevention), thereby distinguishing ideal and ought future selves from one's own perspective or that of significant others. Although teachers may find the L2MSS useful due to its convenience of use and inclusion of the L2LE, which taps into the process of L2 learning, the 2 × 2 model proposed by Papi et al. (2019) excludes this construct and provides a framework for an exclusive understanding future self-guides. The model was developed using rigorous methods (e.g., Papi et al., 2019; Papi & Khajavy, 2021) and has been supported by a growing number of studies that show its relevance to L2 learning motivation (e.g., Huang & Chan, 2024; Li et al., 2025), persistence (Feng & Papi, 2020), emotions (e.g., Jiang & Papi, 2022; Tahmouresi & Papi, 2021), feedback-seeking behaviors (Bondarenko, 2020; Li & Zhang, 2026; Zhang, 2025), and L2 performance and achievement (e.g., Abdi Tabari & Huang, 2025; Papi & Khajavy, 2021; Zhou & Papi, 2023). Taken together, the accumulated evidence suggests that the future of L2 motivation research lies not in abandoning future-self frameworks but in building on and refining them through theoretically and psychometrically stronger models such as the 2 × 2 Model of Future Self-Guides (Papi & Teimouri, 2025).

CRediT authorship contribution statement

Mostafa Papi: Writing – review & editing, Writing – original draft, Validation, Supervision, Data curation, Conceptualization. **Wenting Song:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Hadya Soliman:** Writing – original draft, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Ought-to L2 Self (6 items)
1. I study English because close friends of mine think it is important.
10. If I fail to learn English, I will be letting other people down.
18. I consider learning English important because the people I respect think that I should do it.
26. Studying English is important to me in order to gain the approval of my peers/teachers/family/boss.
34. Learning English is necessary because people surrounding me expect me to do so.
43. Studying English is important to me because other people will respect me more if I have knowledge of English.

Ideal L2 Self (6 items).
9. I can imagine myself speaking English as if I were a native speaker of English.
17. I can imagine myself speaking English with international friends or colleagues.
25. Whenever I think of my future career, I imagine myself using English.
33. I can imagine myself studying in a university where all my courses are taught in English.
41. I can imagine myself writing English e-mails fluently.
51. I can imagine myself living abroad and using English effectively for communicating with the locals.

Attitudes to learning English/L2 Learning Experience (6 items).
54. Do you like the atmosphere of my English classes?
59. Do you find learning English really interesting?
63. Do you think time passes faster while studying English?
67. Do you always look forward to English classes?
71. Would you like to have more English lessons at school?
75. Do you really enjoy learning English?

Intended Effort (6 items).

8. I would like to spend lots of time studying English.
16. I am prepared to expend a lot of effort in learning English.
24. I would like to concentrate on studying English more than any other topic.
32. If an English course were offered in the future, I would like to take it.
40. If my teacher gave the class an optional assignment, I would certainly volunteer to do it.
50. I would like to study English even if I were not required.

Integrativeness (3 items).

56. How much would you like to become similar to the people who speak English?
69. How important do you think learning English is in order to learn more about the culture and art of its speakers?
73. How much do you like the English language?

References

Abdi Tabari, M., & Huang, Z. (2025). Investigating the impacts of task types on the CALF of L2 written production: The moderating role of L2 writing selves. *Reading and Writing*, 1–29. https://doi.org/10.1007/s11145-025-10708-x

Al-Hoorie, A. H., Hiver, P., & In'nami, Y. (2024). The validation crisis in the L2 motivational self system tradition. *Studies in Second Language Acquisition, 46*(2), 307–329. https://doi.org/10.1017/S0272263123000487

Al-Hoorie, A. H., Hiver, P., & In'nami, Y. (2025). Looking beyond the L2 motivational self system: Lessons from two "lost" decades. *Studies in Second Language Acquisition, 47*(4), 1193–1204. https://doi.org/10.1017/S0272263125101319

Bondarenko, A. V. (2020). *Self-efficacy as a generative mechanism for future self-guides and feedback-seeking behavior in language learning* (Publication No. 28153295) [Doctoral dissertation, Florida State University]. ProQuest Dissertations & Theses Global.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (Second Edition). The Guilford Press.

Cho, M. (2021). Regulatory fit effects on the acquisition of lexical stress: A classroom-based study. *Studies in Second Language Acquisition, 43*(5), 1094–1115. https://doi.org/10.1017/S0272263121000334

Cho, M. (2026). Examining regulatory fit effects on L2 writing motivation: Individual regulatory focus and teacher feedback. *International Journal of Applied Linguistics, 36*(2), 1708–1719. https://doi.org/10.1111/ijal.70022

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

Dörnyei, Z. (2006). *Motivation, language attitudes and globalisation: A hungarian perspective, 18*. Multilingual Matters.

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters.

Feng, L., & Papi, M. (2020). Persistence in language learning: The role of grit and future self-guides. *Learning and Individual Differences, 81*, Article 10194. https://doi.org/10.1016/j.lindif.2020.101904

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://www.jstor.org/stable/3151312.

Freitas, A. L., & Higgins, E. T. (2002). Enjoying goal-directed action: The role of regulatory fit. *Psychological Science, 13*(1), 1–6. https://doi.org/10.1111/1467-9280.00401

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy Of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Higgins, E. T. (1987). Self-discrepancy: a theory relating self and affect. *Psychological review, 94*(3), 319. https://doi.org/10.1037/0033-295X.94.3.319

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280–1300. https://doi.org/10.1037/0003-066X.52.12.1280

Higgins, E. T. (2011). *Beyond pleasure and pain: How motivation works*. Oxford University Press.

Huang, H. T., & Chan, H. Y. (2024). Heritage identity and indigenous language learning motivation: A case of indigenous Taiwanese high school students. *The Modern Language Journal, 108*(S1), 127–146. https://doi.org/10.1111/modl.12894

Jiang, C., & Papi, M. (2022). The motivation-anxiety interface in language learning: A regulatory focus perspective. *International Journal of Applied Linguistics, 32*(1), 25–40. https://doi.org/10.1111/ijal.12375

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford Publications.

Lawson, K. M., & Robins, R. W. (2021). Sibling constructs: What are they, why do they matter, and how should you handle them? *Personality and Social Psychology Review, 25*(4), 344–366. https://doi.org/10.1177/10888683211047101

Li, C., Fang, Y., & Derakhshan, A. (2025). Unlocking the interplay among Chinese EFL Learners' L2 motivation, regulatory focus, and language learning achievement: From a regulatory focus theory perspective. *Learning & Motivation, 91*, Article 102141. https://doi.org/10.1016/j.lmot.2025.102141

Li, J., & Zhang, L. J. (2026). Unveiling the antecedents of feedback-seeking behavior in L2 writing: The impact of future L2 writing selves and emotions. *Assessing Writing, 67*, Article 101009. https://doi.org/10.1016/j.asw.2025.101009

McClelland, N., & Larson-Hall, J. (2025). Why you should stop using the ideal L2 self and the L2 motivational self-system to measure motivation (Reaction to Al-Hoorie, Hiver & In'nami, 2024). *Studies in Second Language Acquisition, 47*, 1121–1132. https://doi.org/10.1017/S0272263124000779

Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System, 38*(3), 467–479. https://doi.org/10.1016/j.system.2010.06.011

Papi, M. (2018). Motivation as quality: Regulatory fit effects on incidental vocabulary learning. *Studies in Second Language Acquisition, 40*(4), 707–730. https://doi.org/10.1017/S027226311700033X

Papi, M., & Hiver, P. (2020). Language learning motivation as a complex dynamic system: A global perspective of truth, control, and value. *The Modern Language Journal, 104*(1), 209–232. https://doi.org/10.1111/modl.12624

Papi, M., & Khajavy, G. H. (2021). Motivational mechanisms underlying second language achievement: A regulatory focus perspective. *Language Learning, 71*(2), 537–572. https://doi.org/10.1111/lang.12443

Papi, M., Bondarenko, A. V., Mansouri, S., Feng, L., & Jiang, C. (2019). Rethinking L2 motivation research: The 2× 2 model of L2 self-guides. *Studies in Second Language Acquisition, 41*(2), 337–361. https://doi.org/10.1017/S0272263118000153

Papi, M., & Teimouri, Y. (2025). Manufactured crisis: A response to Al-Hoorie et al. (2024). *Studies in Second Language Acquisition, 47*(4), 1096–1107. https://doi.org/10.1017/S0272263124000494

Papi, M., Zhang, Y., Zhou, Y., Kim, C. J., Mahbodi, M., Eom, M., & Jiang, C. (2025). Regulatory focus and fit effects on task engagement: An experimental study. *Applied Linguistics*. , Article amaf035. https://doi.org/10.1093/applin/amaf035

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods, 25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(1), 1–36. https://doi.org/10.18637/jss.v048.i02

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). Multilingual Matters.

Schoonen, R. (2015). Structural equation modeling in L2 research. *Advancing quantitative methods in second language research* (pp. 213–242). Routledge.

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). Multilingual Matters.

Tahmouresi, S., & Papi, M. (2021). Future selves, enjoyment, and anxiety as predictors of L2 writing achievement. *Journal of Second Language Writing, 53*, Article 100837. https://doi.org/10.1016/j.jslw.2021.100837

Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2016). Discriminant validity testing in marketing: An analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science, 44*(1), 119–134. https://doi.org/10.1007/s11747-015-0455-4

Zhang, Y. (2025). The impact of teacher academic support and L2 writing self on feedback-seeking behavior. *Reading and Writing, 38*(4), 1197–1215. https://doi.org/10.1007/s11145-024-10557-0

Zhou, Y., & Papi, M. (2023). The role of future L2 selves in L2 speech development: A longitudinal study in an instructional setting. System, 119, 103156. https://doi.org/10.1016/j.system.2023.103156.