


CRITICAL COMMENTARY

Validation crisitunity

A response to Al-Hoorie, Hiver, and In'nami (2024)

W. L. Quint Oga-Baldwin 

Waseda University, Tokyo, Japan
Email: quint@waseda.jp

(Received 04 May 2024; Revised 02 September 2024; Accepted 23 September 2024)

Abstract

Al-Hoorie et al. (2024: *Studies in Second Language Acquisition*, 1–23) illuminate a validation crisis within the second language (L2) Motivational Self System (L2MSS), revealing empirical flaws in its current measurement. Their analysis indicates a persistent lack of discriminant validity among the system's constructs, issuing a fundamental challenge in distinguishing the concepts. These findings, echoing previous concerns, underscore a pressing need for theoretical refinement and methodological rigor within the field, leading the authors to advocate for a temporary halt in L2 self-studies to address these issues comprehensively. This commentary discusses the call for a substantive moratorium presented by Al-Hoorie et al. (2024: *Studies in Second Language Acquisition*, 1–23) as a necessary step toward resolving persistent challenges in the field. By highlighting historical issues and suggesting pathways for theoretical diversification and methodological advancement, I aim to foster a productive dialogue on motivational psychology in language learning while ensuring empirical robustness.

Keywords: L2MSS; methodology; open science; response paper; theory

Introduction

Lisa: Did you know that the Chinese use the same word for “crisis” as they do for “opportunity”?

Homer: Yes. Crisitunity!

- *The Simpsons*, Season 6 Episode 11.

Though the above quote is inaccurate with regard to the Chinese language and intended for comedy, the fundamental idea is actually a sound notion for research. Crises can create opportunities for those wise and wary enough to use them. The validation crisis noted in the recent paper by Al-Hoorie et al. (2024) comes out of empirical

and theoretical problems with the second language (L2) Motivational Self System (L2MSS) as it has been measured thus far—and not just on one occasion. Their work indicated that many of the constructs did not adequately differentiate from one another, a failure of what is known as discriminant validity. The authors have raised the alarm to say that failure to adequately distinguish constructs may lead researchers to invalid and unjustified conclusions, requiring a complete bottom-up methodological and theoretical overhaul of the theory before new substantive conclusions can be made. With this overhaul, the crisis can be turned toward an opportunity for improving the general strength and hygiene of research on the psychology of language learning. The validation crisis, better thought of now as a crisis/tunity, requires a fundamental reconsideration of the substantive elements of the L2MSS and a rigorous methodological update and overhaul to match a growing paradigm shift.

A key marker of the need for change can be seen in the attention given to the results and findings. The current article is the third response to this particular article (Henry & Liu, 2024; Papi & Teimouri, 2024). While most empirical work does not generate this level of discourse, the challenge issued by the original article, and the voices raised in nuanced analysis of that challenge (Henry & Liu, 2024) and defense of the status quo (Papi & Teimouri, 2024) are indicative of a Kuhnian paradigm shift (Kuhn, 1962/2012) in the field of the psychology of language learning. To wit, in Kuhn's (1962/2012) description of scientific revolutions, scientific progress is not linear but represented by periods of "normal science" guided by an established paradigm of thought and practices, punctuated by "scientific revolutions" triggered by empirical and theoretical anomalies—like those found by Al-Hoorie et al. (2024) and Henry and Liu (2023, 2024). These revolutionary periods are not merely knowledge advancements but paradigm shifts involving a complete rethinking of the practices and approaches in the field, often facing resistance from the established scientific community due to their disruptive nature. This resistance, mirroring the power struggles seen in political revolutions, highlights the political nature inherent in these scientific upheavals. As it has done before (see Dörnyei & Ryan, 2015), the state of the psychology of language learning has reached the evidentiary, theoretical, and political need for a paradigm shift.

Al-Hoorie et al.'s (2024) work represents the culmination of a critical set of findings, the tipping point where business-as-usual "normal science" has become untenable. Though the opposing voices are engaged in some hair-splitting, angels-on-pinheads reanalysis, and commentary on the work (Papi & Teimouri, 2024), the results follow a pattern of findings and theoretical analyses that indicate that the time for change has come. To date, some work has indicated the need for theoretical overhaul (Henry & Liu, 2023, 2024); other researchers have sought to expand the theory with greater care for obviously missing parts of the model (Papi et al., 2019; Teimouri, 2017; Thorsen et al., 2017); still others have shown further failure to replicate the model as originally hypothesized (Hiver & Al-Hoorie, 2020a). Al-Hoorie et al. (2024) have raised the alarm and warned of the possibility of propagating errors throughout the field of the psychology of language. Given the state of flux in the L2 self-system both substantively and methodologically, the Al-Hoorie et al. (2024) paper rightly calls for a moratorium on substantive studies of L2 selves until these methodological issues can be effectively resolved.

In this commentary, I wish to indicate how Al-Hoorie et al. (2024) have indicated the correct response of a substantive moratorium on the use of the L2 selves, survey some past problems with the literature to highlight the need for greater theoretical and methodological clarity, and present ways to move the field forward that may allow the field of motivational psychology in language learning to continue to include the concept

of an L2 self. Recognize that a substantive moratorium does not call for the topic to be abandoned altogether; rather, there is a need to first get the methods and theory in a workable state before any meaningful educational and psychological findings can be accepted.

Methodological struggles in L2 selves literature

No matter whether we are in agreement or disagreement with every aspect of the interpretation, what Al-Hoorie et al. (2024) have indisputably indicated is that the measurement of the L2 Motivational Self System as practiced to date has several underlying problems. This is not the first time research has indicated fundamental shortcomings; Al-Hoorie's (2018) meta-analysis showed that the key variables of the model (ideal self, ought-to self, and L2 experience) showed wide heterogeneity in their relationship with language achievement. As a meta-analysis, this study should be treated with a fair degree of weight, as it surveyed the entire field up through 2018. More recently, Hiver & Al-Hoorie (2020a) used preregistration to show that one of the largest studies to test this theory (You et al., 2016) failed to replicate.

For the latter study, this was actually unsurprising: many of the fit indices and other standard empirical markers of quality within the original study by You et al. (2016) were close to borderline for the cutoffs used in structural equation modeling. In normal models this would not be problematic, save that given the very large (>10,000) sample size of the population in this initial study, common goodness-of-fit indices are often less sensitive to model misspecifications (Marsh et al., 2004). They were thus likely inflated, and given their borderline nature, might have failed to fit given a smaller sample. Additionally, some of the correlations exceeded .8, indicating collinearity issues above the level noted by Dörnyei (2007) and within the range for lack of differentiation noted by Papi and Teimouri (2024). Thus, though the original study verged on acceptable, it was quite probably misspecified from the start, and the preregistered replication published in 2020 was sorely needed.

In truth, studies using the L2MSS and failing to demonstrate adequate methodological rigor are unfortunately quite prevalent; the highlighted studies by Al-Hoorie et al. (2024) and Hiver & Al-Hoorie (2020a) are exceptional for their use of methods that promote a more parsimonious interpretation of the constructs. While there are debatable aspects of Al-Hoorie et al. (2024), including the choice to use collinearity cutoffs of .6 noted by Papi and Teimouri (2024), these decisions were in fact justified within the literature. I note here that the .6 cutoff came from Dörnyei's (2007) own contention, making the correlations of supposedly different constructs exceeding .8 originally found in You et al. (2016) additionally troubling.

These findings can be seen as extensions compounding errors created early in the research paradigm. The initial studies (Taguchi et al., 2009) creating the instruments most often used in L2MSS studies skipped the creation of a measurement model prior to creating a structural model, a failure of basic modeling practice (Kline, 2023), and hence many of the following studies have struggled to replicate. They further relied heavily on error correlations, a practice often frowned upon in structural equation modeling (Kline, 2023; MacCallum et al., 1992). The problem with correlated errors (likely modeled post hoc through the use of modification indices), is that the model thus relies less on the initial hypothesized theory, and more on potentially unseen confounding factors (see Al-Hoorie & Hiver, 2024, for a full discussion). Though Taguchi et al. (2009) tested their survey instruments across three cultural contexts, each context

had different correlated errors, indicating discriminant validity problems from the outset. The use of highly differing correlated errors in each of the cross-sectional samples without note or explanation of why these residual terms might share variance (e.g., similar wordings in the language used, cultural connections in ideas, proximal placement on the survey) means that the constructs are not as clean or clear as we might desire when describing target qualia.

There were further no measurement invariance tests used to ensure that the constructs were represented by the same structures in all three locations. As Taguchi et al. (2009) is a method validation report, this raises large questions regarding the validity of item structure across localities (DeVellis & Thorpe, 2021; Hirose et al., 2024). Given the differentiations in wordings beyond basic emic considerations for cultural and local understanding, this is somewhat understandable; perhaps the items might not localize in the same way. However, these wording differences were further done without noting theoretical justification or localized validation procedures (Hirose & Oga-Baldwin, 2022), leaving us open to speculate that the wordings chosen were simply based on the judgment of the authors. The models are therefore not directly comparable due to differences in item content and quantity. The constructs thus cannot truly be said to represent the same things across any of the locations studied, which may explain the problems found in L2MSS studies.

Though this study was reported as a book chapter, and thus its inclusion might be overlooked as evidence for general acceptance of weaker fitting models throughout the L2MSS paradigm, this study has also been used as the first instrument creation for widespread empirical testing of the theory, which has led to more mainstream examples of methodological inadequacy. It is cited over 1500 times according to Google Scholar, indicating the potential for misspecified studies to number now into the thousands. The practice of (over)use of modification indices and correlated errors has continued in other well-cited research (e.g., Papi, 2010). Updates to the models and instruments have been largely based on refining and expanding those presented by Taguchi et al. (2009); work by Teimouri (2017; 2018) and Papi and colleagues (2018) has updated some of these ideas, though may also be compounding some of these issues of methodological fuzziness. There have thus been problems with instrumentation and representation of phenomena from the outset, which have propagated down through the research paradigm.

In one of the more egregious examples, Moskovsky et al. (2016) based their instruments on the work by Taguchi et al. (2009), and presented their findings as a structural model while not actually completing one; the work used basic multiple regression without testing structural model fit. It is worth noting here that a composited regression model and a structural model should not be represented in the same fashion. Far worse, the authors failed to recognize that the variables made use of a Likert-type scale rating 1 as “higher” while 5 was “lower” (opposite Taguchi et al., 2009, which read: 1 [Strongly disagree] to 6 [Strongly agree] [p. 90]). They then used these variables to predict International English Language Testing System (IELTS) scores, which would rate in the opposite direction (1 - low, 9 - high). This would explain the negative relationship between the motivational variables and achievement as an erroneous artifact—the study is unique in indicating that lower quantity and poorer quality of motivation lead to better learning. A far more parsimonious interpretation would be that the authors did not appropriately transform their data to align their variables; given the other methodological shortcuts found this seems quite likely.

The authors further exclude a large amount of necessary data (such as how correlations were run without running a structural model or what statistical package

was used for the analyses), while relying on extremely low Cronbach's alpha statistics (.59–.69) to indicate factor unity rather than running proper factor analyses. The failure to account for basic math and methods in the article is quite disappointing given the number of citations the article has generated (246 on Google Scholar at current writing); the authors have failed to respond to repeated inquiries regarding the paper. Its inclusion in one of the top journals in our field indicates an unfortunate acceptance of weak methods when they fit an accepted paradigm (Kuhn, 1962/2012); at least two reviewers and the journal editors had to read and sign off on these results. In an ideal world, this paper would likely be retracted.

In the interest of balance, improvements and theoretical expansions have indeed come through in the intervening years. Some recognition should be given to the work by Papi and colleagues (2018) for updating the scales from the prior models and expanding the theoretical components. While these empirical studies show some improvements, they have not remedied the deeper theoretical problems with the conceptualizations. Indeed, these substantive findings unfortunately stand on shaky theoretical grounds. One major issue in the coming research paradigm will be continuing to ameliorate the statistical and methodological shortcomings.

Theoretical fuzziness in conceptualization

Beyond these existing empirical issues, comments have been made regarding the theoretical consistency of the components of the theory. Henry and Liu (2023) have indicated how the L2 motivational self-system fails to represent a self-system according to most of the common definitions used by social psychologists; their response commentary highlights the jingle-jangle problems and overlaps with the formulations as they currently exist (Henry & Liu, 2024). The L2 motivation self-system may represent a set of phenomena common to language learners, but also may not comprise a complete system of selves in the grander organismic self-regulatory sense (Henry & Liu, 2023). The fundamental philosophical and theoretical conceptualizations of the L2MSS thus stand on shaky ground.

Another aspect of the theory requiring attention is the definitions and terminologies used. With regard to the notion of the *L2 learning experience* (Dörnyei & Ryan, 2015; Dörnyei, 2019), we can note that in jingle fashion, this variable is also sometimes interchangeably called *attitudes to the L2* (Taguchi et al., 2009; Moskovsky et al., 2016, etc.), indicating a terminology issue that has not been resolved within the theoretical community. This variable has been indicated as the best predictor for L2 outcomes: some measures of L2 learning (Al-Hoorie, 2018; Dörnyei, 2019) and the problematic “intended effort” variable. As noted by Al-Hoorie et al. (2024), intended effort is a troubling construct, both because it is quite close in wording to the enjoyment of learning, and further because the gulf between intentions and actions often goes uncrossed (Oga-Baldwin et al., 2019). Unpacking L2 learning experience based on its indicators reveals an even deeper issue. Many of the indicators share similar wording and meaning with measures of intrinsic motivation (Noels et al., 2000; Oga-Baldwin & Nakata, 2017; Ryan & Connell, 1989). Table 1 shares the similarities and overlaps between the L2 learning experience and more widely used measures of intrinsic motivation.

Granted, there are slight differences in the wording, and an empirical comparison is indeed warranted. At the same time, these wordings show a high degree of overlap; perhaps enough to merit serious questioning regarding jangle similarities.

Table 1. Item similarities between L2 learning experience and intrinsic motivation.

L2 learning experience (Taguchi et al., 2009, etc.)	Intrinsic motivation (Oga-Baldwin & Nakata, 2017, translated and adapted from Ryan & Connell, 1989)
I like the atmosphere of my <<English>> class.	<i>Anchor: Why are you studying <<English>>?*</i> Because I like my <<English>> class
I find learning <<English>> really interesting.	Because I am interested in <<English>>
I really enjoy learning <<English>>.	Because I enjoy it

*While English is used in the originals, any language could potentially be inserted here.

Without question, both scales indicate feelings of interest, positivity, and enjoyment in learning a foreign language. As noted, the L2 learning experience is the “Cinderella” variable of the L2MSS (Dörnyei, 2019). At the same time, intrinsic motivation is a well-recognized construct running across many fields of study (Ryan & Deci, 2017). It is perhaps for these among other reasons that Al-Hoorie et al. (2024) proposed self-determination theory as an alternate generalized model for explaining language learning motivation: theoretical parsimony would indicate that if the strongest variable in the L2MSS is substantively the same as intrinsic motivation, it is reasonable to surmise that the model may duplicate effects found in a theory where intrinsic motivation plays a central role. Many of the findings (Al-Hoorie, 2018) may in fact be inadvertently replicating the effects of intrinsic motivation.

Proponents of the L2 selves models have also indicated that the external to internal continuum represented by self-determination theory (Ryan & Deci, 2017) may underlie the structure of the ideal and potential varieties of ought-to L2 selves. Teimouri (2017) nicely charted this continuum, placing the ought-to selves as language-specific versions of external motivations, while noting that the ideal L2 self is similar to intrinsic motivation. From the perspective of parsimony, if these variables can be charted on the same continuum then they are likely best represented by the more universally recognizable formulations which also present a strong set of corollary hypotheses and measurable effects, rather than theoretically tortuous and incomplete constructions with tenuous internal and external validation.

Resolving the jingle, the jangle, and the crisis

Jingle–jangle issues are well-recognized in both education and psychology (Skinner, 2023a, 2023b). Throughout our fields, we find psychometric fuzziness, theoretical overlaps, and yes, even some “thumb-on-the-scale” practices like the overuse of modification indices and correlated errors used to make poorly fitting models look acceptable. The presence of poor practice elsewhere does not excuse us from overlooking it within our own field. On the contrary, it makes appropriate modeling all the more crucial; we can do better.

As Al-Hoorie et al. (2024) have indicated, the scales that have been developed for use with the L2 Motivational Self System show rather extreme problems. Whether or not they are overlapping similar constructs with theoretically unique contributions—a *la* the oft-noted jingle–jangle of self-concept and self-efficacy (Marsh et al., 2019)—is a moot point, given the lack of theoretical consistency with other psychological concepts (Liu & Henry, 2023; Henry & Liu, 2024). The data indicate that where theory states we would expect two or more constructs, we instead often find only one. In looking through the data (kindly made available via open science repositories), the results

show that the factor analytic procedures used are correct. Independent confirmation of the data showed that measures of incremental and absolute fit for some of the most widely used scales do not support these scales as neatly unidimensional; to the contrary, many are theoretically uninterpretable.

Though Papi and Teimouri (2024) provide their own reanalysis of the Al-Hoorie et al. (2024) data, these reinterpretations are quite theoretically tortuous and lack the parsimony of Al-Hoorie et al. (2024). A number of the contentions by Papi and Teimouri (2024) regarding supposed discriminant validity require a post hoc partial correlation of the constructs, seemingly not based on latent factor models. This is not a preferred solution, as latent variables retain all the information of model error terms (Kline, 2023), while mean structures do not. Partial correlations of theoretical constructs thus cannot supersede the more stringent factor analyses carried out by Al-Hoorie et al. (2024); partial correlation analysis alone is not a generally accepted method for determining discriminant validity (Devellis & Thorpe, 2021). Further, these post hoc correlation tests would seem to replicate the “modeling by modification index” problems found in Taguchi et al. (2009). Though some might think the .6 cutoff for construct overlap to be too loose, the failure of the factor analytic procedures to reproduce the hypothesized structure in both You et al. (2016) and Taguchi et al. (2009) confirms the problematic nature of the L2MSS instruments.

The conclusions warrant a substantive moratorium on the use of these scales, as called for by the authors until improved measures are developed. Stopping the propagation of misleading constructs is the heart of this measurement crisis; it is not a frivolous call to panic as indicated by claims of a “manufactured” crisis (Papi & Teimouri, 2024), but a rather a well-reasoned request to improve the quality of research in the psychology of language learning. Note that this substantive moratorium is not a call to cease all research on the topic, but rather a call to first address the theoretical (Henry & Liu, 2023) and methodological problems before attempting to make any unjustifiable practical conclusions. As researchers, this type of moratorium is achieved only by consensus; we as a field can hopefully agree to the conclusion that the L2 Motivational Self System requires an overhaul before we return to a period of “normal science” (Kuhn, 1962/2012). While this may be shocking and frustrating to researchers with projects underway, there are several reasons to retain hope.

1. Al-Hoorie et al. (2024) represent a single (though strong) data point. From the meta-analytic lens (Kline, 2019), projects that are currently in progress (ethics cleared, data gathered, analysis underway), may yet find results contrary to those reported by Al-Hoorie et al. (2024). Some of the substantive papers might be pivoted into validation papers. If the results of these hypothetical studies make use of up-to-date methods with longitudinal designs, appropriate sample sizes, open data sharing, and no post hoc repurposing, they are to be welcomed. At the same time, given that findings incongruous with the theory as developed appear to be the norm (cf. You et al., 2016; Al-Hoorie & Hiver, 2020a), studies indicating parity with theory may well also be the outliers (e.g., Papi et al., 2018). The matter of whether current theory remains valid comes down to an empirical question, one only answerable through good methods with open science.
2. Potential authors of studies that found differing but no less uninterpretable results using the L2MSS can also find success; you did nothing wrong. The baseline instruments were problematic from the start. These instruments can be safely reexamined and reconstructed using approaches similar to those of Al-Hoorie

et al. (2024). Perhaps the work of Al-Hoorie et al. (2024) can provide a theoretical framework for interpretation—and a future chance to preregister models likely to work.

3. Moving forward, the preregistration of models and hypotheses will receive a preference for interpretation (Liu et al., 2023). Ongoing research like that noted above which uses the open science approaches will not be fallible to the same post hoc rationalization. Several preregistered models that perfectly replicate the L2 Self variables as they were originally hypothesized without relying on post hoc modifications to work might be sufficient to overcome the crisis.
4. Caveat: Simply because the instruments and theories do not align when run through factor analytic procedures does not mean that this represents the wrong method. This argument, though sometimes made, is akin to blaming concrete as a building material generally when any house falls down, regardless of the circumstances. Indeed, it may not be ideal in all circumstances, but when used correctly it forms a very stable foundation. At the same time, other methods do indeed exist for determining the dimensionality of constructs (see DeVellis & Thorpe, 2021), and these can also be used. Methods for defining the dimensionality of constructs (e.g., confirmatory composite analysis; Alamer et al., 2024) still rely on similar metrics and heuristics for ensuring discriminant validity, and these metrics and their cutoffs do not sidestep the problems (e.g., multicollinearity and theoretical overlap) indicated by Al-Hoorie et al. (2024) and Henry and Liu (2024).

As implied from the above possible approaches to the crisis, the moratorium/crisis does not necessarily mean a full stop to studies; rather, it is a temporary stop to substantive development until methods are able to catch up. The call does not go out from a perspective of authority, but rather in the same way as a witness, whistleblower, or credible informant raises an alarm noting problems. It is a signal that we as a community can and should come together to ameliorate measurement problems in the field of language learning psychology.

As noted, proper scale development procedures appear to have been overlooked when creating the scales used to define the L2MSS. Though work like that done by Taguchi et al. (2009) is laudable for its attempt at cross-cultural application and diversity of the sample, the reliance on post hoc measures like error correlations indicates that it, at the very least, requires refinements, or more likely needs an overhaul. One way around this will be the use of open science methods, including preregistration of hypotheses and methods (Liu et al., 2023). It is important to remember that properly done replication and validation studies are indeed real research, and deserve equal respect from editors and reviewers.

Moving forward, there is a need for very serious fundamental conceptual, theoretical, and methodological work needed to make the L2 Motivational Self model a reliable measure of motivation, and not simply a “flash in the pan” phenomenon as might be indicated by its rapid rise (Boo et al., 2015; Liu, 2024). Many of the findings to date are unfortunately suspect due to this shaky foundation—given this situation, we cannot be certain that any of the results in this paradigm are trustworthy. For those willing to put in the work, a well-developed program of theoretical exploration and empirical verification could well form several academic careers. However, this will need to be done from first principles. For those seeking theoretical and empirical connection without the need to first lay a foundation, other established traditions may offer a path with fewer obstacles.

Other theoretical avenues

For those wishing to research motivation without the need to reset the fundamentals, there are indeed several examples of lower-hanging fruit. These are offered not as the most correct or most intuitive to language education, but rather available options with the potential for cross-disciplinary appeal.

As a caveat, I come from a specific tradition, and as a statement of bias, I believe that this tradition has much to offer in terms of explanatory power. Philosophically and empirically, self-determination theory offers a clear model that allows for good hypothesis testing while promoting positive well-being and social responsibility (Sugita McEown & Oga-Baldwin, 2019). I fundamentally agree with Al-Hoorie et al. (2024) that channeling resources toward self-determination theory can and will improve the overall landscape of L2 research. Several SDT minitheories remain under-researched (Al-Hoorie et al., 2022), including goal contents minitheory and causality orientations minitheory. Exploring these minitheories can help to increase the explanatory power of the larger metatheory. At the same time, this is by no means the only or optimal set of models that could be adapted for language learning research.

Other theories and constructs abound. Self-efficacy as a construct has received attention in L2 research (Fryer et al., 2024), but it is often researched outside of the grounds under which it was conceived. Though originally developed as part of social cognitive theory (Bandura, 1976; 1986; 1997), it is often operationalized outside this framework, promiscuously inserted into any model that requires a competence component. This is somewhat unfortunate, as social cognitive theory (SCT) offers several powerful elements, not the least of which is a model for how learning occurs through watching others and receiving instruction (Bandura, 1976). As of yet, though self-efficacy has seen several studies in recent years (e.g., Li & Zhang, 2023), SCT more broadly has only been very weakly explored in the L2 realm (Fryer et al., 2024).

At the same time, numerous researchers would prefer to have more language-specific theories. Similar to SCT, Alexander's (2003) model of domain learning (MDL) seeks to explain how learners go from novice to mastery within a specific area. Though a generalized educational theory, it is flexible enough to allow for language specialization (Parkinson & Dinsmore, 2019). Developmental perspectives of this nature complement those of complex dynamic systems (Hiver & Al-Hoorie, 2020b), and when modeled appropriately can show how learners move through different states toward stable outcomes. A hybrid developmental model of both motivation and acquisition could be powerful indeed—and perhaps powerful beyond the language-specific domain.

A dream scenario would be attempts at theoretical comparison, competition, integration, and testing (King & Fryer, 2023). How much variance does SDT's continuum of organismic integration from external to internal regulation explain when run against the L2 selves? Does a discrepancy between the current and ideal self really drive learners toward greater mastery in the MDL, or could that better be explained by a sense of self-efficacy? Which of these constructs might be colliders, changing the strength and direction of effects and outcomes by their presence in a model (Al-Hoorie & Hiver, 2024), which are sibling constructs (Lawson & Robins, 2021), and which are confounders, occupying the same conceptual space? What theoretical variables might actually be parent or second-order factors with multiple effects through the proposed model? Ideally, the current measurement crisis in L2 motivation will find common ground with the ongoing crisis of theoretical diaspora in education and psychology (Skinner, 2023b). How can we find useful and generalizable models that communicate across fields and contexts?

These are empirical questions that can be answered with good designs, good instrumentation, and research hygiene, especially the use of preregistration as the highest standard of such (Liu et al., 2023). At present, the instrumentation that exists for the L2MSS as it was used in Al-Hoorie et al. (2024) has not achieved an acceptable level of trustworthiness, thus leading to the current crisis. Theoretically, the models that have been developed to date require an overhaul before they can be trusted (Henry & Liu, 2023; 2024), lest we send future researchers down the garden path toward disappointment and dead-ends. The jingle-jangle overlaps, intimated by Teimouri (2017), demonstrated in multiple uses of terminology (cf. Dörnyei & Ryan, 2015; Taguchi et al., 2009), and indicated in highly similar item wordings (see above), must be addressed through appropriate methods. Issues of discriminant validity must be handled through factor analysis and other multivariate latent dimensional scaling procedures (Devellis & Thorpe, 2021; Kline, 2023).

In the long term, dominant research paradigms are chosen through a complex combination of both evidence and political will (Kuhn, 1962/2012). Despite widespread political acceptance, the evidentiary aspect has clearly been lacking in the L2MSS. Whether it is in SDT, SCT, MDL, L2MSS, or a hybridized version of any or all of the above, all these models and theories require a preponderance of evidence for sustainable acceptance. The crisis offers researchers the chance to build the coming paradigm of open science and model competition.

It should be noted that this research paradigm is not alone in calls for a substantive pause for reasons of psychometric inadequacy. Given the state of the evidence in recent years, researchers have noted the need for reevaluation of several very popular theories and models, including self-regulated learning (Dinsmore & Fryer, 2023) and even basic psychological need satisfaction and frustration as it is used in SDT (Murphy et al., 2023). Reassessments and refinements are a part of science.

The process of science often requires the reestablishment of fundamentals that overturn the work of large figures in the field. William Thomson, known more famously as Lord Kelvin, originator of the Kelvin measurement scale and giant of nineteenth-century science wielded a famously outsized influence on science through his role in the Royal Society (Thompson, 1910). Despite being publicly wrong about topics as varied as the age of the Earth and Darwinian evolution (Badash, 1989), the existence of X-rays, and the possibility of heavier-than-air flight (Thompson, 1910), his legacy remains even today. Between 1870 and 1890, his vortex theory of atomic structure (Thomson, 1867) was the dominant paradigm in chemistry and atomic physics but was ultimately overturned using new instrumentation, opening opportunities for later researchers such as Rutherford, Bohr, Röntgen, and Einstein to develop the physical and chemical sciences that have built the modern world. Improved measurement resolved the issues: empirical demonstration and improved theory were synthesized to become the politically accepted reality. Each well-measured but theoretically puzzling result led to new pathways for inquiry that need not be beholden to demonstrably outmoded methods and theories; the psychology of language learning is faced now with a similar shift.

For researchers with projects underway, this may mean a need to pivot. At the same time, if these researchers are able to make these same scales work using up-to-date psychometric approaches, their L2MSS projects are likely to reap rewards. For others, the need for fundamental restructuring will move them toward theories with improved parsimony. Until the crisis is resolved, any substantive work on the nature of the L2 Motivational Self System is just as likely to be wrong, and thus overturned by competing models. Taking advantage of this crisis will allow researchers to make their mark

and carve out new theoretical and methodological directions in the coming research paradigm.

Data availability statement. The datasets generated during and/or analyzed during the current study are not publicly available due to privacy concerns but are available from the corresponding author upon reasonable request.

Acknowledgment. This work was supported by a Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (23H00647).

Competing interest. The author declares none.

References

- Alamer, A., Schubert, F., & Henseler, J. (2024). When and how to use confirmatory composite analysis (CCA) in second language research. *Studies in Second Language Acquisition*, 1–20. <https://doi.org/10.1017/S0272263124000020>
- Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32(8), 10–14.
- Al-Hoorie, A. H. (2018). The L2 motivational self system: A meta-analysis. *Studies in Second Language Learning and Teaching*, 8(4), 721–754. <https://doi.org/10.14746/sslt.2018.8.4.2>
- Al-Hoorie, A. H., Oga-Baldwin, W. L. Q., Hiver, P., & Vitta, J. P. (2022). Self-determination mini-theories in second language learning: A systematic review of three decades of research. *Language Teaching Research*, 136216882211026. <https://doi.org/10.1177/13621688221102686>
- 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*, 1–23. <https://doi.org/10.1017/S0272263123000487>
- Al-Hoorie, A. H. & Hiver, P. (2024). Causality and ability beliefs: An introduction to confounders and colliders. *Studies in Second Language Learning & Teaching*.
- Badash, L. (1989). The age of the Earth debate. *Scientific American*, 261(2/August), 90–97. <https://www.jstor.org/stable/10.2307/24987366>
- Bandura, A. (1976). *Social learning theory*. Prentice-Hall.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bandura, A. (1997). *Self efficacy: The exercise of control*. W. H. Freeman.
- 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>
- DeVellis, R. F. & Thorpe, C. T. (2021). *Scale development: Theory and Applications* (5th Edition). Sage.
- Dinsmore, D. L., & Fryer, L. K. (2023). Critical thinking and its relation to strategic processing. *Educational Psychology Review*, 35. <https://doi.org/10.1007/s10648-023-09755-z>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Dörnyei, Z. and Ryan, S. (2015). *The Psychology of the Language Learner Revisited*. Routledge.
- Dörnyei, Z. (2019). Towards a better understanding of the L2 Learning Experience, the Cinderella of the L2 Motivational Self System. *Studies in Second Language Learning and Teaching*, 9(1), 19–30. <https://doi.org/10.14746/sslt.2019.9.1.2>
- Fryer, L. K., Li, C., Guo, Z., Liang, L., & Zhong, Y. (2024). Self-efficacy's role within learning a new language during formal education: Systematic review, critical evaluation of past research and paths forward for research and practice. *Studies in Second Language Learning and Teaching*.
- Henry, A., & Liu, M. (2023). Can L2 motivation be modelled as a self-system? A critical assessment. *System*, 119, 103158. <https://doi.org/10.1016/j.system.2023.103158>
- Henry, A., & Liu, M. (2024). Jingle-jangle fallacies in L2 Motivational Self System research: A response to Al-Hoorie et al. (2024). *Applied Linguistics*, 45(4), 738–746. <https://doi.org/10.1093/applin/amae041>
- Hirosawa, E., & Oga-Baldwin, W. L. Q. (2022). Surveys and questionnaires with young language learners. In Y. G. Butler & B. H. Huang, *Research Methods for Understanding Child Second Language Development* (pp. 33–48). Routledge. <https://doi.org/10.4324/9780367815783-3>

- Hirosawa, E., Kono, Y., & Oga-Baldwin, W. L. Q. (2024). The structure of ability beliefs in EFL classrooms: A cross-theoretical analysis bridging self-efficacy and perceived competence needs satisfaction. *System*, 124, 103383. <https://doi.org/10.1016/j.system.2024.103383>
- Hiver, P., & Al-Hoorie, A. H. (2020a). Reexamining the role of vision in second language motivation: A preregistered conceptual replication of You, Dörnyei, and Csizér (2016). *Language Learning*, 70(1), 48–102. <https://doi.org/10.1111/lang.12371>
- Hiver, P., & Al-Hoorie, A. H. (2020b). *Research methods for complexity theory in applied linguistics*. Multilingual Matters.
- King, R., & Fryer, L. K. (2023). Hybridizing motivational strains: How integrative theoretical models can advance the motivational sciences. *Educational Psychology Review*, 36. <https://doi.org/10.1007/s10648-024-09850-9>
- Kline, R. (2019). *Becoming a behavioral sciences researcher* (2nd Ed.). Guilford Publications.
- Kline, R. (2023). *Principles and practice of structural equation modeling* (5th Ed.). Guilford Publications.
- Kuhn, T. (1962/2012). *The structure of scientific revolutions*. University of Chicago Press.
- 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*.
- Li, C., & Zhang, L. J. (2023). The development of accuracy and fluency in second language (L2) speaking related to self-efficacy through online scaffolding: A latent growth curve modeling analysis. *Journal of Psycholinguistic Research*, 1–25. <https://doi.org/10.1007/s10936-023-09950-7>
- Liu, M., Chong, S. W., Marsden, E., McManus, K., Morgan-Short, K., Al-Hoorie, A. H., Plonsky, L., Bolibaug, C., Hiver, P., Winke, P., Huensch, A., & Hui, B. (2023). Open scholarship in applied linguistics: What, why, and how. *Language Teaching*, 56(3), 432–437. <https://doi.org/10.1017/S0261444822000349>
- Liu, M. (2024). Mapping the landscape of research on the L2 motivational self system theory (2005–2021): A bibliometric and text network analysis. *System*, 120, 103180. <https://doi.org/10.1016/j.system.2023.103180>
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111, 490–504.
- Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9, 275–300. <https://doi.org/10.1037/1082-989X.9.3.275>
- Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, 111(2), 331–353. <https://doi.org/10.1037/edu0000281>
- Moskovsky, C., Assulaïmani, 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>
- Murphy, B. A., Watts, A. L., Baker, Z. G., Don, B. P., Jolink, T. A., & Algoe, S. B. (2023). The Basic Psychological Need Satisfaction and Frustration Scales probably do not validly measure need frustration. *Psychological Assessment*, 35(2), 127–139. <https://doi.org/10.1037/pas0001193>
- Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. L. (2000). Why are you learning a second language? Motivational orientations and selfdetermination theory. *Language Learning*, 50(1), 57–85. <https://doi.org/10.1111/0023-8333.00111>
- Oga-Baldwin, W. L. Q., & Nakata, Y. (2017). Engagement, gender, and motivation: A predictive model for Japanese young language learners. *System*, 65, 151–163. <https://doi.org/10.1016/j.system.2017.01.011>
- Oga-Baldwin, W. L. Q., Fryer, L. K., & Larson-Hall, J. (2019). The critical role of the individual in language education: New directions from the learning sciences. *System*, 102118. <https://doi.org/10.1016/j.system.2019.102118>
- 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., 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. and Teimouri, Y. (2024). Manufactured Crisis: A Response to Al-Hoorie et al. (2024). *Studies in Second Language Acquisition*.

- Parkinson, M. M., & Dinsmore, D. L. (2019). Understanding the developmental trajectory of second language acquisition and foreign language teaching and learning using the Model of Domain Learning. *System*, 86, 102125. <https://doi.org/10.1016/j.system.2019.102125>
- Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57(5), 749–761. <https://doi.org/10.1037/0022-3514.57.5.749>
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Skinner, E. A. (2023a). Four guideposts toward an integrated model of academic motivation: Motivational resilience, academic identity, complex social ecologies, and development. *Educational Psychology Review*, 35(3), 80. <https://doi.org/10.1007/s10648-023-09790-w>
- Skinner, E. A. (2023b). Is academic motivation a tree trunk, a fan, a wall, a rope, a snake, or a spear? No, it's an elephant and it's on fire. In M. Bong, J. Reeve, & S.-I. Kim (Eds.), *Motivation science: Controversies and insights* (pp. 371–379). Oxford University Press.
- Sugita McEown, M., & Oga-Baldwin, W. L. Q. (2019). Self-determination for all language learners: New applications for formal language education. *System*, 86, 102124. <https://doi.org/10.1016/j.system.2019.102124>
- Taguchi, T., Magid, M., & Papi, M. (2009) The L2 Motivational Self System among Japanese, Chinese, and Iranian learners of English: A comparative study. In: Dörnyei, Z. and Ushioda, E., Eds., *Motivation, Language Identity and the L2 Self* (pp. 66–97), Multilingual Matters.
- Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, 39, 681–709. <https://doi.org/10.1017/S0272263116000243>
- Thomson, W. (1867). On vortex atoms, *Proceedings of the Royal Society of Edinburgh*, 6, 94–105. https://zapatopi.net/kelvin/papers/on_vortex_atoms.html
- Thompson, S. (1910). *The Life of William Thomson, Baron Kelvin of Largs*. MacMillan and Co., Limited. <https://archive.org/details/lifeofwillthom02thomrich/page/n7/mode/2up>
- 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*, 23(5), 584–600. <https://doi.org/10.1080/13670050.2017.1388356>
- You, C., Dörnyei, Z., & Csizér, K. (2016). Motivation, vision, and gender: A survey of learners of English in China. *Language Learning*, 66, 94–123. <https://doi.org/10.1111/lang.12140>