Issues in Defining and Measuring Mindsets and Other Psychological Phenomena

Paul Collett
Shimonoseki City University

Reference Data:

Collett, P. (2024). Issues in defining and measuring mindsets and other psychological phenomena. In B. Lacy, R. P. Lege, & P. Ferguson (Eds.), *Growth Mindset in Language Education*. JALT. https://doi.org/10.37546/JALTPCP2023-18

In research of such psychological phenomena as mindsets it has been argued that researchers are prone to various methodological and theoretical missteps that can weaken the validity and reliability of research conclusions. This paper serves as a brief overview of some of the issues that have been facing researchers in psychology, and which are seen to be contributing to a theory crisis in psychology. Problems concerning defining and measuring psychological phenomena, the validity of operationalized constructs, and the implications of methodological missteps on research outcomes are discussed. These are all issues arising from an orthodoxy in research which adheres to questionable empiricist methodologies grounded in logical positivist conceptions of the research process. As it is likely the same state of affairs will arise when researching these phenomena in relation to language learning, the aim is to explicate some suggestions into how methodologically-sound research may be conducted and proceed.

マインドセットのような心理現象の研究において、研究者は様々な方法論的・理論的誤謬を犯しがちであり、それが研究結論の妥当性や信頼性を弱めていると論じられてきた。本稿は、心理学の研究者が直面し、心理学における理論の危機を助長していると考えられている問題のいくつかを簡単に概観したものである。心理学的現象の定義と測定に関する問題、運用化された構成概念の妥当性、そして方法論的な誤謬が研究成果に与える影響について論じている。これらはすべて、研究プロセスに関する論理実証主義的な概念に基づく、疑わしい経験主義的な方法論に固執する研究の正統性から生じる問題である。言語学習に関連するこれらの現象を研究する際にも、同じような状況が生じる可能性が高いため、方法論的に健全な研究をどのように実施し、進めていくべきかについて、いくつかの示唆を述べることを目的とする。

JALT2023, with its theme of "Growth Mindset in Language Education" was in some ways a coming of age for the concept of mindsets in relation to foreign or second language learning in Japan. There has been a limited amount of research into the concept in the Japanese language learning and teaching context over the last decade and a half, probably the most noticeable being that of Mercer and Ryan (2010; Ryan & Mercer, 2012). As the concept of mindsets reaches a larger audience and likely spurs further research, it is an opportune time to consider some of the issues involved in researching this and other psychological constructs, and to take stock of methodological discussions into the nature of the theorised phenomena themselves. This paper provides an overview of the methodological and theoretical issues confronting researchers at the intersection of psychology and foreign language learning. Problems concerning defining and measuring psychological phenomena, the validity of operationalized constructs, and the implications of methodological missteps on research outcomes are discussed. These are all issues arising from an orthodoxy in research which adheres to questionable empiricist methodologies grounded in logical positivist conceptions of the research process. An alternative approach based around a critical realist account of how research should best proceed is introduced as a viable way forward to help understand and account for the place of psychological phenomenon in language learning.

Defining the Phenomena

It is not uncommon for research in a number of areas of foreign language learning to draw on concepts from psychology. Motivation, learner strategies, learner and teacher beliefs, autonomy, amongst other fields, all draw on concepts from psychology, adopting them to account for the language learning process. However, in doing so, researchers should be aware that they may be relying on approaches that lack the methodological rigour required to detect true effects of their studies, and that they may be perpetuating methodological and epistemological missteps that are at the base of the theory crisis (Oberauer & Lewandowsky, 2019) said to be afflicting psychological research at the current time.

Many of the psychological phenomena that have been brought into language learning research are unobservable and thus not easily quantifiable or qualifiable. The common approach to measure them, at least in quantitative research, is to operationalise the phenomenon of interest as a construct in a manner that allows it to be measured. One of the first concerns here relates to this operationalism. Garrison (2022) provided the following critiques of this process. Firstly, it limits the scope of discovery by reducing observation to that which the observer is interested in, potentially meaning other data that may be related to the phenomena under study may be overlooked. Next, operationalism leads to identification of the phenomena to be studied with the tool being used to measure it; so, for example, language ability becomes something that is measured by standardized English tests (e.g., IELTS, TOEFL, TOEIC). Due to its origins within empirical approaches to psychology, operationalism reduces phenomena to observable data, and the data becomes confounded with the phenomena. Here for example, a student's identity as a language learner may be codified by their placement on an assessment test.

Another problem is that in many cases, the validity of such operationalised constructs may not be apparent, so it is unclear exactly what is being measured. Within psychological methodology, Flake and Fried (2020) identified a failure to clearly define the construct being tested as one of a number of questionable measurement practices (QMP) that limit the value of research. Al-Hoorie and Vitta (2018), in a review of statistical reporting in a number of second-language journals, found that validity was often unreported. If it is not clear what is being measured, it means that the same construct could be operationalised under different names or measured using different techniques, creating confusion about what exactly the construct is and how to analyse it. Al-Hoorie (2018) argued that this is the case with research into language self systems, where there is "redundancy & conceptual clutter" (p. 738) over the exact definition of the concept of self. Al-Hoorie, et. al. (2024) went as far as to claim there is a crisis concerning the validity of the methods used in researching this field, something that is likely to be found across research into other psychological constructs due to similarities in the methodological approaches employed.

A slightly different occurrence that can also potentially waste research resources and time and add to conceptual ambiguity is when attempts are made to introduce a new concept into research where an extant alternative may well account for the same findings. Jahedizadeh and Al-Hoorie (2021) suggested this may be the case with directional motivational currents, which may ultimately best be conceptualised as the more well-researched phenomenon of flow (Dörnyei, et al., 2016). This issue has also risen in the debate regarding language learning strategies. Multiple definitions have attempted to explain what learning strategies are—Oxford (2017) catalogued 33 different definitions. However, rather than providing any sort of conceptual clarity, this multitude has contributed to "…chaos and to some condemnations of the field" (Oxford, 2017, p. 1). In these cases where even the definition cannot be agreed upon, it is difficult to formulate a theory to account for findings as theories are potentially underdetermined (Eronen & Bringmann, 2021). Underdetermination means that the evidence available is not sufficient to be able to establish the best theory to account for patterns of data and phenomena observed.

Measurement

Once a phenomenon has been operationalized the next step is to consider how the phenomenon can be effectively measured. In the case of language mindset research (e.g., Lou & Noels, 2016, 2017), data collection is often through the creation and use of self-reporting tools such as questionnaires which are intended to measure theorised dimensions of the mindset. Once the results are collected, they are usually correlated against some measure operationalised to represent another aspect of learner motivation such as an aspect of self-regulation or self-efficacy. These results are then often correlated with outcomes on measures of achievement or capability to show how the mindset is related to the academic domain under study. Al-Hoorie, et al. (2021) critiqued this kind of approach when used to access motivational interventions as "the questionnaire curse" (p. 141) and pointed out that as observational data it is not very helpful in determining the effectiveness of that which is being measured.

Often, researchers will use Likert scale questionnaires when collecting quantitative data, and then report and carry out analysis on the means of the responses to each item making up the scale. Whether this is justifiable or not involves a long-running debate (Brown, 2016; Carifio & Perla, 2007); needless to say, whatever methodological stance one takes, it should be clearly outlined and appropriate analyses should be applied for the type of measurement being carried out. It is important that consumers of the research can understand the research process and rationale behind it if the research is to practically contribute to the advance of knowledge.

Flake and Fried (2020), in their enumeration of multiple QMP that may be problematic for research, pointed to a lack of transparency and poor reporting practices in the research process as an important factor which can reduce the value of research findings. They concluded that much research into various psychological constructs

has been limited by inappropriate measurement choices and insufficient attention to methodology, making for low-quality measurement instruments. This leads to low quality data collection, which makes phenomena hard to detect, a point brought up by Eronen and Bringmann (2021). They also discussed the problem of "fat-handedness" of interventions when doing research into psychological processes. Trying to change one particular variable in the process of research to assess its impact on phenomena is almost always going to have a knock-on effect on other parts of the psychological system. As these processes are not observable, and due to problems with operationalism, it is not clear to what extent these other changes will be apparent nor how much they will account for any data measured. Here, the question becomes one of the validity of data, and how this can be assured.

In studies related to language learning, cross-cultural issues should also be addressed. When studying psychological phenomena, as with many other variables, it is not a simple matter to import a methodology used in one culture into a new one. Measurement tools need to be tested and adjusted to fit the conditions of the testing environment, which usually requires an exploratory process to generate a suitable data collection instrument before confirmatory research can take place. More broadly, there is the issue of etic and emic perspectives in research. An etic perspective is akin to an "outsider looking in" to a sociocultural milieu. The assumption here is that there are universal or shared phenomena across cultures which allows a researcher who is not part of the group being studied to understand what is being observed or measured. An emic perspective is an insider's viewpoint, that is, the idea that one must be part of and cognizant of the norms of a group to be able to fully account for data collected. If the researchers are adapting a framework developed to explain phenomena in a different cultural context to that in which the research is carried out, can they then make any claims to gaining a socioculturally valid understanding of their subjects' responses, or can the results only be understood in terms of the milieu in which the framework was originally developed? Given findings such as Henrich et al. (2010) regarding the potential limitation of outcomes of behavioural science research to a particular subset of tertiary-educated North Americans, the question of the validity of applying such research in different cultural contexts seems particularly troubling.

Making sense of the data measured is obviously a major consideration, which leads into the appropriateness of data analytic techniques. Factor analysis (FA) is a common statistical tool used in the analysis of psychological constructs. Through FA, various factors that are assumed to account for the structure of the data found can be generated, which can then help with modelling or theorising the nature of the construct. FA can be either exploratory (EFA) or confirmatory (CFA), with each having different roles in the research process (Loewen & Gonulal, 2015; Field, et al., 2012). Unfortunately, it is not uncommon to find this method used incorrectly. For example, EFA is first used to posit a number of factors that may explain the data, with a CFA then applied to the same data in an attempt to show that the factors do explain the data to x degree of likelihood. This approach is problematic: "Refrain from performing exploratory and confirmatory analyses on the same dataset as this yields high danger of overfitting, in particular in smaller datasets." (Fokkema & Greiff, 2017, p. 401).

Overfitting is when there are exaggerated estimates of parameters, with a model fitting too closely to the data and, thus, being non-generalisable. Furthermore, "It is not logical to obtain a good-fitting factor structure using EFA and then seek to confirm that structure using CFA with the same dataset, doing so capitalizes on sample-specific, chance relationships and in no way verifies the EFA findings. Unfortunately, there are published examples of this poor practice" (Flora, & Flake, 2017, p. 85). In fact, highly rated journals in the language learning research field have also published research articles making this exact methodological mistake. For example, cases can be found in volumes of *System* and *Studies in Second Language Acquisition* from recent years. Although the authors of these papers may have had a defensible reason for their methodology, this is not clearly articulated. This is problematic in that it suggests an EFA followed by CFA on the same data is generally appropriate when in most cases it should not be used. A better approach is to carry out the EFA on one sample and then do the CFA on a second, different sample, or to randomly split a sample into two and carry out the EFA and CFA on the two separate groups (Hair, et. al., 2019). Either option brings with it the added cost or research burden of having a sample large enough to be able to detect or verify the factors.

The issue of sample size arises in relation to other types of data analysis, too. When using inferential statistics to analyse data, best practices dictate that researchers should always carry out a power analysis before engaging in the research (Plonsky, 2013). This allows the researcher to see what sample size would be required to measure an effect size of x with, for example, the significance level set at $p = 0.05$ and test power of 0.8 or more. As it stands, the effect sizes for many of the phenomena studied in social psychology tend to be small. A magnitude of $d = 0.4$ is given as a minimum benchmark to judge the effectiveness of educational effects (Jahedizadeh & Al-Hoorie, 2021), which, according to Cohen's (1988, 1992) convention, is a small to medium effect. Evidence suggests that phenomena measured in social psychology have small to medium effects of somewhere between $r = 0.18$ to 0.3, with a mean $r = 0.2$ (Schäfer & Schwarz, 2019). Note

that although these are different types of effect size indices, $r = 0.2$ is equivalent to $d = 0.42$ (Watt, 2021); the average effect is at the minimum level for what can be judged as effective. This should not be surprising given that many of these phenomena are one part of interrelated complex systems with multiple effects influencing outcomes in relation to the item being measured.

What this means in practical terms is that large samples are required to detect these kinds of phenomena. For example, if you were planning on comparing two independent groups to assess if there was any difference in some data you had measured, you could use a *t*-test to analyze the data. Assuming, based on your knowledge of the topic, that you are expecting an effect size of $d = 0.4$, you would require a sample size of 100 subjects in each group to get a result that allows you to act on the null hypothesis with acceptable levels of probability that you will not be committing a Type I or Type II error. The former of these is when you reject your null hypothesis when it is true, the latter when you do not reject a false null hypothesis.

Another popular inferential statistic used in research into psychological phenomena is Analysis of Variance (ANOVA); some studies suggest it is the most common inferential statistical method used in research in the field of language acquisition (Plonsky, 2013; Khany & Tazik, 2018). When carrying out an ANOVA, once the initial *F* score has been calculated, it is then usually necessary to do further testing to find where any difference detected between groups lies. It is standard practice, and advisable, to carry out post-hoc tests (Brown, 2016). However, one limitation here is that they lack precision and may increase the risks of Type I errors. A better approach is to use planned contrasts. Here, you are testing the groups of interest against one another, rather than testing the relationships amongst all groups; in effect testing a directional hypothesis. This gives your tests more precision (Field, et al., 2012; Lindstromberg, 2016b). While planned contrasts are somewhat more complex to carry out than the standard post-hoc tests often reported, it is probably good practice to use them in place of (or along with) the latter so you can have more confidence in analytical outcomes of data.

Mistaken decisions in the data analytic stage of research can lead to failures to provide quantitatively sound outcomes. This is an issue related to limited understanding and improper application of statistical methods, which seems common across foreign/second language acquisition research fields (Al-Hoorie & Vitta, 2018; Gonulal, 2020; Lindstromberg, 2016a). Unintentional misuse of methods, underpowered studies, and a lack of robustness of statistical tests are ongoing problems here (Collett, 2021; 2022). On a broader level is a misplaced confidence in the role of statistical hypothesis testing, a potentially inappropriate methodology if the necessary testing conditions have not been met (Scheel et al., 2021). It may also be that conclusions drawn from hypothesis testing are limited in explanatory scope, providing little substantive value (Borsboom et al., 2021). Discussions of these problems and recommendations for alternative approaches have been ongoing for years (Collett, 2022). The issues are seen as being behind the low replicability of research results (Open Science Collaboration, 2015) and contributing to a wider theory crisis (Eronen & Bringmann, 2021).

Improving the Research Process

All these issues point to the problematic nature of research into psychological factors in language learning. However, they are the same concerns that researchers within psychology are facing, with multiple solutions being put forward, some of which have been touched on above. Overall, perhaps the best suggestions that can be offered are as follows.

Little has been said about qualitative research here, but Ushioda (2020) provided an informed perspective on current approaches based on in-depth analysis of narrative data collected from subjects. This data could include interviews, learner and teacher journals, and other *in situ* records of the learning-teaching process. The importance of understanding a person-in-context approach is also stressed, recognising and working to interpret the influences of people's experiences on their perspectives.

Lou and Noels (2019) have suggested as such in their outline of a newly proposed Language Mindset Meaning System (LMMS) to explain what they believe is the central role of mindset in language learning motivation. They acknowledge the need to locate their research within a broad sociocultural context. Lou and Noels (2019) pointed out: "...conceptualizing and assessing language mindsets should not simply change a few wordings from the general mindset scale...we need to thoroughly validate new measurements as findings in one domain or setting may not translate directly to others." (p. 555). Research into mindsets, as well as other psychological areas, should be a collective process requiring collaboration across different fields. Furthermore, Lou and Noels (2019) acknowledged their LMMS as part of a complex dynamic system, reflecting a trend of research into psychological phenomena and language learning. Ushioda (2020) pointed out that qualitative methods are ideally suited for research in these contexts.

Mixed-methods research is a good alternative approach to adopt, where both quantitative and qualitative methods are combined without prioritising the results of one approach over another. Tremblay (2020) brings us up to date with some developments for quantitative analysis that can benefit researchers in language learning,

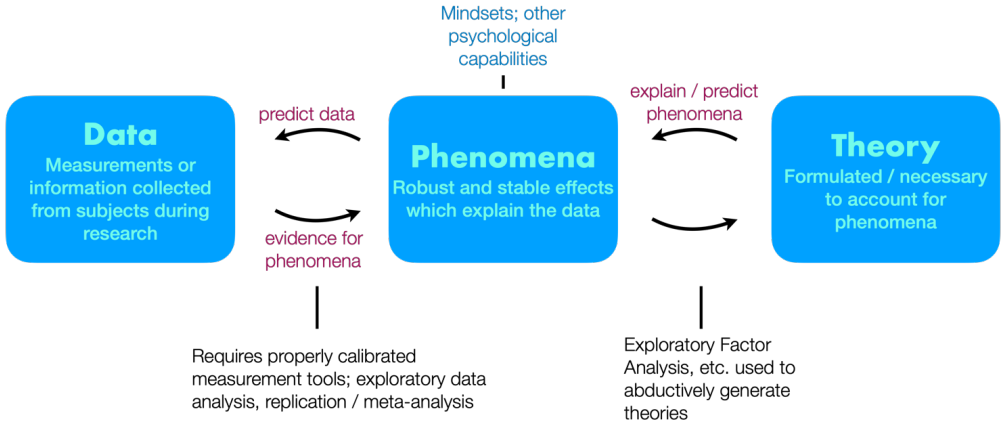
such as advances in structural equation modeling and mixed modeling. His perspective is that these methods can help build richer models to explain the relationships between phenomena. This could help play a part in the exploratory stage of research, with qualitative analysis then being used to provide a deeper understanding of the phenomena and to help uncover causal mechanisms at work.

In terms of a lack of transparency and poor reporting of measurement practices as outlined by Flake and Fried (2020), they presented a number of questions researchers should address to ensure a more rigorous research process. This is in part echoed by Eronen and Bringmann (2021) who pointed out that "conceptual clarification and construct validation should be seen as an important and valuable points (sic) of research and validation should be taken to be an ongoing process instead of just a hurdle that needs to be crossed" (p. 785). In some instances, some of the problems discussed may be due to restrictions on word counts / page length when publishing research, or editorial suggestions to leave out what may seem irrelevant to the general report. The "Open Science" initiative—an open-access program designed to increase the transparency, honesty, and robustness of research—is being promoted as a solution to this issue and has received support from the Japanese government (Open Science Framework, n.d.; Cabinet Office, n.d.). Another possibility is for publications to provide online space for appendices of supplementary materials, datasets, scripts created for data analysis, and other materials that allows for replication and verification of the results, as is happening across different domains to varying degrees now.

Many of the problems outlined in this paper are the result of the ongoing influence of logical positivism and empiricism on the scientific process. Ideally, moving away from a hypothetico-deductive approach to research, and the baggage this brings from positivist and empiricist epistemologies, may be the best stance. Adopting methods following abductive logic is one recommendation (Haig, 2005, 2013, 2018). Abduction, also known as inference to the best explanation, is the method of drawing conclusions which are the likeliest possible explanation for some condition based on what you already know. Chirkov (2016, as cited in Chirkov & Anderson, 2018, p. 741) defined it as "a way of thinking that occurs when a researcher starts with a problem and empirical evidence regarding it and then suggests a hypothesis of why this problem exists and how the empirical evidence can be explained". Haig (2013, 2018) argued for such methodology involving the detection of phenomena from data combined with abduction to generate explanatory theories that can account for the phenomena. Such a realist approach, wherein the emphasis is on the existence of attributes (or phenomena) that have causal properties, and which relies on predominantly qualitative research methods, seems to sidestep issues of construct validity and operationalism (Borsboom, et al., 2004; Chirkov & Anderson, 2018b; Haig & Evers, 2016). Borsboom et al. (2021) provided a clear overview of this theory construction process. Briefly, the theory construction process distinguishes among data, phenomena, and theory. *Data* is the measurement or information collected from subjects during research. *Phenomena* are the robust, stable effects that explain the data. *Theory* necessary to account for the phenomena is then formulated. This should be able to both explain and predict phenomena, whilst data should be predictable by phenomena. A graphical representation of this model, adapted from Borsboom et al. (2021), is presented in Figure 1.

Figure 1

A Model of Theory Development (based on Borsboom et al. (2021)



A worked example of such an approach to research, contrasting empirical and realist methodologies, can be found in Chirkov and Anderson (2018a, 2018b). As a theoretically grounded practical account of how research into psychological phenomena may best advance, I strongly recommend taking time to read these papers.

Conclusion

It is difficult to say whether second language/foreign language research focusing on psychological variables is anywhere near the stage of the theory crisis seen to be

prevalent in mainstream psychological research. At least some well-researched areas have come in for criticism due to the lack of validation of research tools (Al-Hoorie, et. al., 2024). If researchers analyzing the role of a phenomenon like mindsets in language learning follow the same lines of research, the outcomes could be problematic. More generally, continuing to adhere to a research methodology that attempts to cast statistical regularities as causality is regressive. Researchers are embedded in complex dynamic systems to as much a degree as the subjects of our study. To ensure advancement of knowledge in language learning research and theory, critical appraisal of the nature of these systems, and the practices embedded in them would be laudable. Research into mindsets of language learners is still a relatively open field of study. Being aware of potential problems in methodology is an important first step in research into such psychological phenomena and their influence on language learning.

Bio Data

Paul Collett works in the Organization for Liberal Arts and Teaching at Shimonoseki City University. His educational background is in psychology, with a focus on methodological and epistemological issues in research.

References

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/10.1111/ijal.12416

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., Kim, T.-Y., & De Costa, P. I. (2021). The identity crisis in language motivation research. *Journal of Language and Social Psychology, 40*(1), 136-153. https://doi.org/10.1177/0261927X20964507

Al-Hoorie, A. H., & Vitta, J. P. (2018). The seven sins of L2 research: A review of 30 journals' statistical quality and their CiteScore, SJR, SNIP, JCR Impact Factors. *Language Teaching Research, 23*(6), 727-744. https://doi.org/10.1177/1362168818767191

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science, 16*(4), 756–766. https://doi.org/10.1177/1745691620969647

Brown, J. D. (2016). *Statistics corner: Questions and answers about language testing statistics.* JALT.

Cabinet Office. (n.d). 国際的動向を踏まえたオープンサイエンスに関する検討会 [Study group on open science in light of international trends.] Cabinet Office, Government of Japan. Retrieved January 7, 2024, from https://www8.cao.go.jp/cstp/sonota/openscience/

Carifio, J. & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences, 3*(3), 106-116. https://doi.org/10.3844/jssp.2007.106.116

Chirkov, V., & Anderson, J. (2018a). Statistical positivism versus critical scientific realism. A comparison of two paradigms for motivation research: Part 1. A philosophical and empirical analysis of statistical positivism. *Theory & Psychology*, *28*(6), 712-736. https://doi.org/10.1177/0959354318804670

Chirkov, V., & Anderson, J. (2018b). Statistical positivism versus critical scientific realism. A comparison of two paradigms for motivation research: Part 2. A philosophical and empirical analysis of critical scientific realism. *Theory & Psychology*, *28*(6), 737-756. https://doi.org/10.1177/0959354318816829

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cohen J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Collett, P. (2021). Graphical data analysis. In P. Clements, R. Derrah, & P. Ferguson (Eds.), *Communities of teachers & learners*. JALT. https://doi.org/10.37546/JALTPCP2020-01

Collett, P. (2022). An online application for exploratory and explanatory data analysis. In P. Ferguson, & R. Derrah (Eds.), *Reflections and new perspectives*. JALT. https://doi.org/10.37546/JALTPCP2021-26

Dörnyei, Z., Henry, A., & Muir, C. (2016). *Motivational currents in language learning: Frameworks for focused interventions*. Routledge.

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science, 16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications.

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science 49*(2), 78-88. https://psycnet.apa.org/doi/10.1037/cbs0000069

Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble: Overfitting in the assessment of internal structure and some editorial thoughts on it. *European Journal of Psychological Assessment 33*(6), 399-402. https://psycnet.apa.org/doi/10.1027/1015-5759/a000460

Garrison, M. J. (2022). Critical research methodology: A realist approach. In K. J. Saltman & N. Nguyen (Eds.), *Handbook of critical approaches to politics and policy of education* (1st ed.). Routledge. https://doi.org/10.4324/9781003145356

Gonulal, T. (2020). Statistical knowledge and training in second language acquisition: The case of doctoral students. *International Journal of Applied Linguistics, 171*(1), 62-89. https://doi.org/10.1075/itl.18031.gon

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods, 10*(4), 371–388. https://doi.org/10.1037/1082-989X.10.4.371

Haig, B. D. (2013). Detecting psychological phenomena: Taking bottom-up research seriously. *The American Journal of Psychology, 126*(2), 135–153. https://doi.org/10.5406/amerjpsyc.126.2.0135

Haig, B. D. (2018). *The philosophy of quantitative methods*. Oxford University Press.

Haig, B. D., & Evers, C. W. (2016). *Realist inquiry in social science.* Sage.

Hair, J.F., L.D.S. Gabriel, M., da Silva, D. & Braga Junior, S. (2019). Development and validation of attitudes measurement scales: Fundamental and practical aspects. *RAUSP Management Journal, 54*(4), 490-507. https://doi.org/10.1108/RAUSP-05-2019-0098

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral And Brain Sciences, 33*, 61-135. https://doi.org/10.1017/S0140525X0999152X

Jahedizadeh, S., & Al-Hoorie, A. H. (2021). Directed motivational currents: A systematic review. *Studies in Second Language Learning and Teaching 11*(4), 517-541. https://dx.doi.org/10.14746/sslt.2021.11.4.3

Khany, R., & Tazik, K. (2018). Levels of statistical use in applied linguistics research articles: From 1986 to 2015. *Journal of Quantitative Linguistics, 26*(1), 48-65. https://doi.org/10.1080/09296174.2017.1421498

Lindstromberg, S. (2016a). Inferential statistics in language teaching research: A review and ways forward. *Language Teaching Research, 20*(6), 741–768. https://doi.org/10.1177/1362168816649979

Lindstromberg, S. (2016b). Guidelines, recommendations, and supplementary discussion. *Language Teaching Research, 20*(6). Retrieved from https://journals.sagepub.com/doi/suppl/10.1177/1362168816649979/suppl_file/10.1177_1362168816651895.pdf

Loewen, S., & Gonulal, T. (2015). Exploratory factor analysis and principal components analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 182-212). Routledge.

Lou, N. M., & Noels, K. A. (2016). Changing language mindsets: Implications for goal orientations and responses to failure in and outside the second language classroom. *Contemporary Educational Psychology*, *46*, 22-33. https://doi.org/10.1016/j.cedpsych.2016.03.004

Lou, N. M., & Noels, K. A. (2017). Measuring language mindsets and modeling their relations with goal orientations and emotional and behavioral responses in failure situations. *The Modern Language Journal, 101*(1), 214-243. https://doi.org/10.1111/modl.12380

Lou, N. M., & Noels, K. A. (2019). Language mindset, meaning-making, and motivation. In M. Lamb, K. Csizér, A. Henry, & S Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 537-559). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3_26

Mercer, S., & Ryan, S. (2010). A mindset for EFL: Learners' beliefs about the role of natural talent. *ELT Journal, 64*(4), 436-444.

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review, 26*, 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

Open Science Framework. (n.d.). OSF Home. Retrieved on January 7, 2024, from https://osf.io

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*. https://doi.org/10.1126/science.aac4716

Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). Routledge.

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*, 655–687. https://doi.org/10.1017/S0272263113000399

Ryan, S., & Mercer, S. (2012). Language learning mindsets across cultural settings: English learners in Austria and Japan. *OnCUE Journal, 6*(1), 6-22.

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.00813

Scheel, A. N., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science, 16*(4), 744-755. https://doi.org/10.1177/1745691620966795

Tremblay, P. F. (2020). Quantitative methods in second language learning motivation research: Gardner's contributions and some new developments. In A. H. Al-Hoorie & P. D. MacIntyre (Eds.), *Contemporary language motivation theory: 60 years since Gardner and Lambert (1959)* (pp. 212-232). Multilingual Matters.

Ushioda, E. (2020). Researching L2 motivation: Re-evaluating the role of qualitative inquiry, or the 'wine and conversation' approach. In A. H. Al-Hoorie & P. D. MacIntyre (Eds.), *Contemporary language motivation theory: 60 years since Gardner and Lambert (1959)* (pp. 194-211). Multilingual Matters.

Watt, R. (2021). Effect sizes: Relating Cohen's d to r. Retrieved January 6, 2024, from https://campuspress.stir.ac.uk/doingpsychstats/2021/02/15/effect-sizes-relating-cohens-d-to-r/