

Foreword

Trevor A. Holster
trevholster@gmail.com
TEVAL SIG Publication Chair

Welcome to the first issue of *Shiken* for 2015. In this issue, Koizumi et al. investigated regression to the mean (RTM) and the standard error of difference (SED), issues which are surprisingly underrepresented in TESOL literature. Although Koizumi et al. and Swinton (1983) demonstrated RTM using TOEFL scores, this problem will affect any pretest-posttest comparisons, whether for research studies or for monitoring student learning. As both Swinton and Koizumi et al. made clear, RTM can lead to misinterpretation of test results, invalidating research findings and compromising instructional decisions, so the lack of awareness of the problem is worrisome. Swinton's study, published as an official ETS Research Report, recommended using two pretests to estimate and correct for RTM, so Koizumi et al.'s guidance on how to estimate RTM when only a single pretest is available provides a much more practical approach for the situation that most teachers and program administrators face.

Smiley, who will be familiar to many TEVAL members as the coordinator of the JALT Materials Writers SIG, provided an account of the difficulties that teachers face in developing skills in test analysis. Smiley's use of Microsoft Excel to conduct classical test theory (CTT) analysis was highly commendable and the guidance provided by books such as Brown's (2005) *Testing in Language Programs* is more than sufficient for the needs of teachers who need to analyze classroom tests. However, Smiley needed to criterion reference test questions against curriculum objectives and textbook content, a purpose for which Rasch analysis is ideally suited (Brown & Hudson, 2002). Linacre's (2014) Winsteps software package provides for quick and detailed Rasch analysis of overall test performance and individual items, but Smiley also cautioned that novices may be discouraged by the steepness of the learning curve involved in learning Rasch analysis.

J.W. Lake and I responded to Smiley's article by highlighting two key features of the Rasch model: the Wright map and the assumption that all items discriminate equally. Our aim was not to provide groundbreaking new insights, but rather to demonstrate that Rasch analysis can provide information to guide instructional decisions that is not easily available from CTT analysis and that Rasch results can be presented in graphical forms that are conceptually simple enough that novices can interpret them without requiring extensive technical training.

Finally, J.D. Brown's regular *Statistics Corner* column reviewed the range of techniques and analyses that have been used in the testing of intercultural pragmatics ability. Pragmatics, which deals with the relationship between context and meaning, is crucial to language proficiency, and thus to assessment, evidenced by the growing body of research on its assessment documented by Brown. Hopefully the inclusion of pragmatic features in assessment will result in positive washback, where textbooks and classroom instruction reflect the testing of intercultural pragmatics. One point that stands out about Brown's review is the increasing sophistication of the analyses used in testing intercultural pragmatics ability, which is evidence of the complex nature of the interaction between language and context. In particular, the increasing use of Facets analysis to account for rater effects (see McNamara, 1996, for an accessible introduction) raises questions about how to incorporate pragmatics into classroom assessments because teachers frequently act as interlocutors and/or raters. Given that pragmatics is concerned with what is appropriate in different contexts and when faced with different interlocutors, the elicitation of pragmatics performances in a classroom by a teacher raises questions of how to interpret the results, i.e. the construct validity of the assessment. The problematic nature of authenticity in classroom contexts is well recognized (see van Lier, 1996, for example). Facets analysis, which isolates contextual variables as

facets of a performance, can address some of these concerns, but the complexity of the analysis often makes the findings incomprehensible to non-specialists, as Smiley's article in this issue reported. This doesn't preclude positive washback from tests of intercultural pragmatics, but it does raise questions about what degree of assessment literacy teachers need for positive washback to occur.

The TEVAL SIG has been working for many years to make technical issues more accessible to classroom teachers through J.D. Brown's *Statistics Corner*, Jim Sick's series of articles on Rasch analysis, and Tim Newfield's articles on assessment literacy, but the articles by Smiley and Brown are important reminders of the need for workshops and introductory articles aimed at novice language testers. The JALT Pan-SIG2015 conference will be held in Kobe on the weekend of 16-17 May, 2015. Many of our officers and members will be attending, so this is an excellent opportunity to see the work of TEVAL SIG members and the members of other JALT SIGs, and to raise any questions or concerns about testing and assessment. We look forward to seeing you at Pan-SIG2015.

References

- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment (New Ed.)*. New York: McGraw-Hill.
- Brown, J. D., & Hudson, T. D. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University Press.
- Linacre, J. M. (2014). Winsteps (Version 3.81.0). Retrieved from <http://www.winsteps.com>
- McNamara, T. F. (1996). *Measuring second language performance*. Harlow: Pearson Education.
- Swinton, S. S. (1983). *A manual for assessing language growth in instructional settings*. Princeton: Educational Testing Service.
- van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. Harlow: Pearson.

Call for Papers

Shiken is seeking submissions for publication in the November 2015 issue. Submissions received by 1 September, 2015 will be considered, although earlier submission is strongly encouraged to allow time for review and revision. *Shiken* aims to publish articles concerning language assessment issues relevant to classroom practitioners and language program administrators. This includes, but is not limited to, research papers, replication studies, review articles, informed opinion pieces, technical advice articles, and qualitative descriptions of classroom testing issues. Article length should reflect the purpose of the article. Short, focused articles that are accessible to non-specialists are preferred and we reserve the right to edit submissions for relevance and length. Research papers should range from 4000 to 8000 words, but longer articles are acceptable provided they are clearly focused and relevant. Novice researchers are encouraged to submit, but should aim for short papers that address a single research question. Longer articles will generally only be accepted from established researchers with publication experience. Opinion pieces should be of 3000 words or less and focus on a single main issue. Many aspects of language testing draw justified criticism and we welcome articles critical of existing practices, but authors must provide evidence to support any empirical claims made. Isolated anecdotes or claims based on "commonsense" are not a sufficient evidential basis for publication.