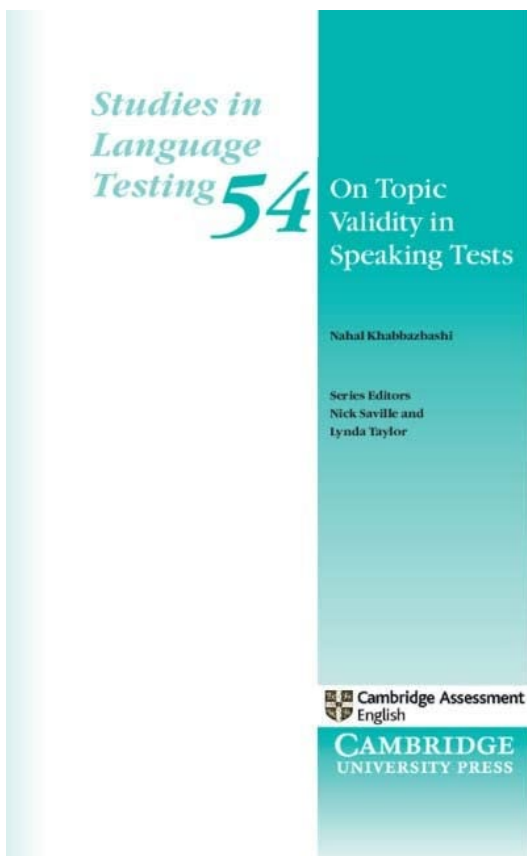




Book Review

On Topic Validity in Speaking Tests, by Lynda Tylor and Nick Saville (Eds.), Cambridge University Press, 2021, 233 pp., \$67.75 U.S. (paperback), ISBN 9781009102490

The discussion concerning how topics and background knowledge of topics (BK) influence test takers' speaking



performance is a fundamental issue that deserves more attention in research. With numerous high-stakes tests being topic-based, it is of great importance to consider the extent that topic differences, familiarity and BK that could affect the assessment of test-taker's speaking ability. These considerations would allow assessment on speaking abilities to be more trustworthy. This book by Khabbazzashi (2021), titled *On Topic Validity in Speaking Test*, provides comprehensive knowledge on the role of topics and BK in topic-based speaking tests. The book takes the IELTS speaking test (IST) as the context and adopts the socio-cognitive framework (SCF) to link test development and validation to the social, cognitive, and evaluative dimensions of language use (O'Sullivan & Weir, 2011). It is divided into seven chapters, each of which is informed based on the author's doctoral dissertation.

This book opens with an introductory chapter that provides a rationale for why topics and BK are under-researched, despite their importance in eliciting speech in performance-based assessments. The author starts by introducing factors that may also affect ratings in second language (L2) assessments and two plausible factors (i.e., task complexities and test-taker characteristics) in relation to the topics and BK. This chapter also

introduces SCF as the framework used to present the validity evidence of how topics and BK affect ratings in the IST. In chapter 2, the author explores seven areas to gain a deeper understanding of how each area considers topics and BK in topic-based speaking tests: (1) second language acquisition (SLA), (2) task-based language assessment (TBLA), (3) reading comprehension, (4) listening comprehension, (5) writing performance, (6) speaking performance, and (7) IELTS-related research.

In chapter 3, the author introduces the additional third factor (i.e., raters) closely associated with topics and BK. She outlines two important reasons for considering this rater factor in the analysis (i.e., the allure of perfect



agreement and the intricacies of the rating process). She then introduces a many-facet Rasch measurement (MFRM) model, a statistical tool designed to handle the influence of these two pivotal (topics and BK) and three additional factors (task, test takers, and raters) on human ratings.

Moving on to chapter 4, the author delineates the mixed methods approach by explaining how she collected the data involving how Iranian learners of English were selected and asked to perform two speaking tests from the pre-selected ten topics. A total of 8 questionnaire items for each topic was developed, and a C-test that distinguishes BK from general language proficiency was completed. The validation measures of the two main instruments, i.e., the C-test and the questionnaire are reported before moving to the finding chapters. In this context, the topics of the IST were examined to ensure they matched the varying levels of BK. Two additional instruments, such as IELTS rating scale was explained to understand the used rating categories and observation checklists were described to evaluate the distribution of the speaking task difficulties. In analyzing the data, a figure of two different MFRM analyses (i.e., a four-facet MFRM to understand the influence of topics and a five-facet MFRM to examine the influence of topics and BK) is visualized. With regard to the qualitative approach, thematic analysis is selected to analyze the transcripts of test takers' speaking performance and those of interviews with raters.

In chapters 5 and 6, the author reports the findings to collect validity evidence of topics and BK from the two research strands. Quantitative analysis forms the focus of chapter 5. The chapter begins with a discussion of the four-facet MFRM analysis, showing that the test takers demonstrated a wide range of IELTS speaking ability levels, the raters exercised different severity levels, and the tasks measured different criteria with increasing difficulty levels. To examine the influence of topics, two analyses were observed: (1) the MFRM analysis of three task types (A, B, C), which revealed that only task C had a practical influence on the performance scores; (2) the analysis of the function of the topics which informs that topic may not have any influence on the performance scores. She then investigated the impact of BK through two different analyses: (1) five-facet MFRM analysis and (2) three-facet MFRM combined with regression analyses, suggesting that BK accounts for the second most important predictor of speaking performance after the proficiency levels. Given this, the test takers with high BK levels benefit more than those with low levels of BK.

In chapter 6, the author examines three types of qualitative data (e.g., rater's interview responses, test-takers' speech, and test-takers' questionnaire responses) for the qualitative research part. This chapter indicates that topic unfamiliarity and lack of BK can pose challenges for both test takers and raters during the assessment process. She further explains that this problem results from "local validity" (p.167), which refers to difficulties arising from socio-cultural and religious backgrounds that some test takers would not have knowledge of. The book concludes by synthesizing the common ground between the two research strands in chapter 7.

Overall, the book offers a rich contribution that fills the knowledge gap on L2 topic-based speaking assessment validity. One key strength is that the book is built around key readings, such as Weir's (2005) and O'Sullivan and Weir's (2011) socio-cognitive framework, offering readers an understanding of how different sources of validity are fitted together to present a great example of a unified approach to validity in L2 speaking assessments. Given that the book is structured with a research-based style, the book is a great source for readers interested in the validity of topic-based tests. The second strength is its use of a widely recognized test, IELTS, that assessment scholars from different parts of the world would have knowledge about. This would empower the scholars to comprehend the information from the first chapter and develop direct applications that fit their contexts. The third strength lies in its clear description of the methodology section in chapter 4, which provides readers with a clear step-by-step procedure for undertaking a validation study of topic-based tests. Given this depth of information, readers would have a bigger picture, which is important for replicating this validity research. The final strength of this book is the comprehensive discussions of the MFRM analyses covered in chapter 5 and the thematic analysis in chapter 6. Findings in chapter 5 present sufficient evidence that emphasizes the role of BK as a second predictor of L2 speaking performance and as a factor affecting task difficulty levels, providing washback evidence for L2 learners to focus on self-background-knowledge improvement. In chapter 6, the author effectively demonstrates how a qualitative method under thematic analysis complements quantitative findings, highlighting how topic

unfamiliarity and lack of BK significantly impact test takers and raters. With these qualities, this book meets the writing standards of a high-quality dissertation and merits publication in Cambridge's Studies in Language Testing (SILT) series.

While the book provides valuable insights on validity studies for topic-based tests, a few limitations are worth mentioning. First, the book appeals to a targeted readership, i.e., some language assessment scholars (p. x), posing challenges to average readers who need to gain knowledge of the methodologies and technical terms, such as Rasch, item response theory for comprehending the book's content. Using more accessible terms would engage a greater number of readers at varying levels of understanding. The second point relates to the literature review of background knowledge. Given the use of inference-based methods to establish the operational definition of BK that does not provide a clear concept of BK, it would be valuable if the author defined the terms; otherwise, it would be risky for a research-based book. Finally, despite having a significant number of participants, their profiles are insufficient to generalize the results. Given that the IELTS speaking tasks were employed in this research, the participants involved should not be restricted to a homogenous group but instead represent a heterogenous group with diverse cultural and linguistic backgrounds.

Overall, the book is an excellent contribution to the validity research literature, with most chapters providing a remarkable illustration regarding the use of the SCF and MFRM in the investigation of topic-based tests. Built on a topic that most researchers seem to neglect, the book has undoubtedly motivated readers to consider the roles of topics and BK in topic-based tests. To conclude, the book provides a great use for language assessment scholars who would like to use the SCF as the framework for validity research.

References

- O' Sullivan, B., & Weir, C. J. (2011). Language testing and validation. In O'Sullivan, B (Ed.), *Language testing: Theory & practice* (pp.3–13). Palgrave.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.

Widya Ratna Kusumaningrum
Iowa State University, USA
widya@iastate.edu

(Received June 20, 2023; Revised December 20, 2023; Accepted March 10, 2024)