

# Reviews





**Wojciech Malec,**  
***Developing Web-based Language Tests.* Lublin:**  
**Wydawnictwo KUL, 2018,**  
**ISBN 978-83-8061-641-7, 439 pages**

The book by Wojciech Malec entitled *Developing Web-based Language Tests* makes a positive impression at first glance because it is well over four hundred pages. Secondly, and more importantly, its topic—foreign language testing, which constitutes a difficult area for teachers inexperienced and experienced alike—suggests that it could be a very valuable resource. Additionally, his presentation of the technological advancements which complement a thorough theoretical background to the subject covers in the greatest details the process of FL test construction, its administration and analysis of the final product. Particular attention is paid to FL tests’ reliability, a complex construct which is explained here in a reader-friendly manner.

The volume consists of two main parts organised into eight chapters, a bibliography consisting of 544 entries, three appendices, and the subject index. The book starts with an *Introduction*, which presents the theme, defines the constructs basic for the topic of language testing, and elaborates on the structure of the volume. The presentation is coherent, logical, and comprehensive in detail. The first part of the book Principles of Assessment (pp. 25–199) constitutes the theoretical background to foreign language testing, whereas the second part Web-based Testing (pp. 201–377) presents the practical aspects of testing, the major part of which is a detailed demonstration of the online programme WebClass, designed by the author. This part shows what role modern computer technologies can play in testing language achievement. The book closes with Conclusions (pp. 379–383). Such a structure gives evidence of what the main objectives of this publication are. Half of the book is a thorough overview of testing as an essential aspect of FL teaching, which is interpreted on the basis

of a very extensive literature in the field. The other half places emphasis on the deployment of information and communication technologies (ICT) in FL assessment. Below there are more detailed comments on the contents of the book and their assessment.

Chapter 1 Test Development (pp. 27–64) is devoted to a detailed description of the FL test construction process, which is based on the very broad selection of sources the Author refers to, making use of older ones such as works by Bachman, Palmer (1996), Niemierko (1999), and more recent publications, just to mention Carr (2011), Lane, Raymon, and Haladyn (2016) among many others. The Author emphasizes here the cyclical nature of constructing a test referring to each of the stages in detail but paying special attention to the first stage (the first component). This stage embraces the design of the test in terms of its context and objectives of assessment, requirements, form and its basis in the profile of the prospective testees. It also comments on the structure and content of the test, as well as on the scale of assessment for each test task.

In Chapter 2 Test Evaluation (pp. 65–124), the Reader will find a presentation of different models of test evaluation such as a model of usefulness by Bachman and Palmer (1996) or the argumentative validation model of Kane (1992–2013). The Author compares both models demonstrating their similarities and differences and taking a critical perspective on both. The Author discusses in a lot of detail the qualities of a good language test, referring among other things to its practicality, authenticity, and validity. However, it is test reliability as the major quality that is rightly assumed by the Author to require the most thorough discussion. Here, the reader will find two different methods of test reliability evaluation, the classical and the generalizability theory. It is the latter one that the Author discusses by means of a presentation of one-facet crossed designs, focusing on relative and absolute errors, coefficients and test length, Phi lambda, standard error of measurement and GT-1 calculator. The chapter also offers a comment on the evaluation of decision consistency (threshold loss agreement, square-error loss agreement) and additionally, the validation procedures of classroom-based tests.

He emphasizes that the analysis of test results in the generalizability theory is a complex process, although there are programmes which can assist testers and facilitate this process. It is worth mentioning at this point is that the Author created his own programme, independent of the WebClass presented later in the book, which is available for FL teachers for use in analysing their test results.

Chapter 3 Test Items (pp. 125–166) demonstrates the process of test item construction, the actions and decisions the teacher has to make in the process of preparing the first version of a test as well as the final one. It is a logical, coherently presented and complete set of test formats. It discusses the selected-response items (multiple choice, binary choice, multiple response, multiple-

choice cloze, matching, etc.) as well as limited-production items (gap-filling, cloze and C-test, gapped sentences, transformation, sentence writing, error correction, etc.). It is not only a first-class description but also a critical assessment of individual test formats. This part of the book proposes useful guidelines for teachers but it is also a necessary introduction to the test formats used in Part II of the book—in the practical discussion of the WebClass platform designed for FL test construction.

Chapter 4 Item Analysis (pp. 167–199) is an extensive comment on how to assess the correctness of test items by means of selected statistical procedures in two types of assessment: norm-referenced testing and criterion-referenced testing. The procedures refer to such parameters as item facility, discrimination, and distractor evaluation. The Author presents a critical view of these procedures, which leads him to his own ideas on how to modify them.

It is another example of how a thorough knowledge of testing issues, critical thinking and a creative approach can lead to new solutions. We have the best example of the above in the practical Part II of the book, presenting a new programme for test construction and its evaluation.

The practical part starts with Chapter 5 Technology in Language Testing (pp. 203–248) and it is a perfect example of the Author's belief in the power of modern technologies, which—when used appropriately, in the educational contexts and specifically in foreign language instruction—will not only facilitate but also improve this process, making it both valid and reliable. In this context of FL testing, the programme described here offers teachers not only help in test construction but also in the evaluation of its reliability at different stages—at the beginning but also at the final stage of the analysis of test results. The platform offers many possibilities, such as class enrolment, peer correction, adding and editing documents, performance report for a given group, feedback options or the possibility of archiving the test for later use (which is not without value for the teacher), among many others. Malec discusses the strengths of the platform but he is also aware of its limitation. It is important to emphasize that WebClass was created on the basis of scholarly theories and findings presented in the first part of the book, thus it testifies to certain scholarly standards. This chapter also presents the platform and its functioning, itemising its four modules—administration, communication, materials, assessment. Additionally, a commentary is offered on teaching and testing with WebClass. In other words, the chapter is a detailed discussion of what the platform has on offer for FL teachers. The following chapters, that is, Chapter 6 Test Design and Production on WebClass (pp. 249–304), Chapter 7 Test Use and Evaluation on WebClass (pp. 305–350) and Chapter 8 Administration Mode (pp. 351–378) continue in terms of technicalities an extremely detailed presentation of the platform, which I will not comment on here as they are purely technical instructions on operating the software/online platform.

Chapter 8 Administration Mode Effects (pp. 351–378) is the only empirical section in the book, which reports on the study comparing the results of two equivalent language tests administered in different forms, that is, a traditional paper-based test (*PBT*) and web-based test (*WBT*). The Author carried out extensive analysis of both tests and their results, measuring, for example, their reliability or item facility. He also looks in this study at single test results in relation to individual learner differences such as, for instance, learners' experience in computer work and its impact on the test results. The analysis demonstrates the comparability of the results in each of the measured aspects—for this reason, the Author recommends the use of both forms of testing.

In Conclusions (pp. 379–384), the Author goes back to the different aspects of FL testing discussed earlier in the theoretical part to emphasize that each of the principles of testing are implemented in his programme. This offers some assurance of its academic validity. Despite the fact that the Author is an ardent promoter of IC technologies in FL testing, he also sees the value in traditional paper-based testing. Although the platform was designed with FL teachers in mind, it can equally well be implemented in other areas of measurements, where analysis of reliability is at issue.

To recapitulate, what I find most interesting and valuable in the book is the fact that it is a text which is really well-read in the literature of language testing, both in relation to the background texts presenting traditional views and especially those which relate to IC technology. Both were used in creating an innovative and extremely useful platform for language testing. Thus, the pragmatic value of the book is well worth noting. It presents not only a useful tool for FL teachers and discusses the platform's merits, but also gives detailed and user-friendly instruction on how to use it. Additionally, teachers will find here an extensive presentation of statistical measures that can be used to assess FL test reliability and ways of analysing results. Having sensitized FL teachers (the prospective readers of this book) to the issues in test construction, analysis and assessment, the Author shares with them his knowledge on test construction at its various different stages. He might also have commented on the most commonly committed errors by teachers in test construction, which is often the weakest dimension of FL teachers' professional competence. As to the practical part of the book, the Author presented only one short empirical study to demonstrate the value of WebClass. Maybe another such examples could be a study of how the Author himself uses the platform in his own daily didactic practice and a discussion of its advantages as measured by selected research tools (also disadvantages and problems). Such examples would constitute a good way of promoting this useful tool. The book concludes with a short subject index. It is a pity that the Author did not also include an authors' index. As to the formal side of the book, although it is written in good academic English, the topic is

not presented hermetically and prospective readers will enjoy its reader-friendly form of expression and also the tidiness of the publication.

All in all, I can fully recommend this book by Wojciech Malec to all FL teachers as well as to researchers on e-learning approaches to foreign language instruction. It is a good example of how knowledge, didactic practice, and passion can result in something as creative and useful as the WebClass platform. I hope that this review may contribute to its success.



<https://orcid.org/0000-0003-0626-0703>

*Danuta Gabryś-Barker*  
*University of Silesia in Katowice, Poland*

