

Günther Sigott (Ed./Hrsg.)

## Language Testing in Austria: Taking Stock

Sprachtesten in Österreich:  
Eine Bestandsaufnahme



In the area of the educational standards, context validity, theory-based validity, scoring validity and standardsetting, rating validity and benchmarking, criterion-related validity as well as consequential validity have been addressed to different extents. *Context validity* and *theory-based validity* is addressed in a study by Siller and Kipman, who examine contextual and cognitive factors that account for differences in item difficulty in the E8 reading test. Pibal, Sigott and Cesnik have studied the role that language errors play in the rating of writing performances ← 16 | 17 → in the E8 writing baseline test, thus shedding light on criteria for correctness that are consciously or subconsciously applied by the raters. *Problems of scoring validity and standardsetting* are addressed in a study by Sigott and Cesnik, who, after describing approaches to standardsetting for the receptive skills in English, focus particularly on the discrepancies between empirical item difficulty and the difficulty estimates made by a standardsetting panel and ways of dealing with these discrepancies without asking panellists to negotiate agreement. Freunberger, Breit and Illetschko report on a study in the area of *rating validity and benchmarking*. They describe the rater training for the writing part of the German test at the end of primary school and document changes in rater behaviour from the training phase to operational testing with particular focus on rater severity and rater agreement. In particular, they identify texts with regard to which the raters tend to disagree and suggest research into features which are characteristic of such controversial texts. Cesnik focuses on an