

# **Annual Technical Report for ACCESS for ELLs Online English Language Proficiency Test**

Series 602, 2023-2024 Administration  
Annual Technical Report No. 20A

## **References**

Prepared by Center for Applied Linguistics  
Language Assessment Division  
Psychometrics and Quantitative Research Team  
June 2025

## References

- Allen, N. L., Carlson, J. E., & Zalanak, C. A. (1999). The NAEP 1996 technical report. National Center for Education Statistics.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.
- American Institutes of Research. (2018). *ELPA21 Technical Report, Part I—Summative assessment*.
- Andrich, D. A. (1978). A rating scale formulation for ordered response categories. *Psychometrika*, 43, 561–573.
- Bachman L. F., & Palmer A. S. (2010). *Language assessments in practice*. Oxford University Press.
- Baker, F. B., & Kim, S.-H. (2017). *The basics of item response theory using R*. Springer.
- Brennan, R. (2004). *Linking with equivalent group or single group design (LEGS)* (Version 2.0) [Computer software]. Center for Advanced Studies in Measurement and Assessment.
- Center for Applied Linguistics. (2016). *ACCESS for ELLs Series 400 listening and reading scale maintenance: Technical brief*.
- Center for Applied Linguistics. (2017). *ACCESS for ELLs 2.0 speaking and writing score scale reconstruction: Technical brief*.
- Center for Applied Linguistics. (2019). *Maintaining the ACCESS for ELLs online writing scale: Preparations for the Series 501 redesign: Technical brief*.
- Chapman, M., Chuang, P., Bitterman, T., & Elliott, H. (2024, August). *Development of a new WIDA writing scoring rubric for grades 1–12* [Technical Report]. WIDA, Wisconsin Center for Education Research, and the Board of Regents of the University of Wisconsin System. <https://wida.wisc.edu/sites/default/files/resource/Technical-Report-Development-New-WIDA-Writing-Scoring-Rubric-Grades-1-12.pdf>
- Chuang, P.-L. (2024, April). *Validating a new writing scoring scale using multi-faceted Rasch analyses* [Technical report]. WIDA, Wisconsin Center for Education Research, and the Board of Regents of the University of Wisconsin System.

<https://wida.wisc.edu/resources/validating-new-writing-scoring-scale-using-multi-faceted-rasch-analyses>

- Cook, H. G., & MacGregor, D. (2017). *The ACCESS for ELLs 2.0 2016 standard-setting study* [Technical Report]. WIDA and the Board of Regents of the University of Wisconsin System.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Engelhard, G., Jr., & Wind, S. A. (2018). *Invariant measurement with raters and rating scales: Rasch models for rater-mediated assessments*. Routledge/Taylor & Francis Group.
- Ercikan, K., & Julian, M. (2002). Classification accuracy of assigning student performance to proficiency levels: Guidelines for assessment design. *Applied Measurement in Education*, 15(3), 269–294.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 105–146). Macmillan.
- Gottlieb, M. (2004). *English language proficiency standards for English language learners in kindergarten through grade 12: Framework for large-scale state and classroom assessment*. WIDA and the Board of Regents of the University of Wisconsin System.
- Jones, P. E., & Smith, R. W. (2006, April). *Item parameter drift in certification exams and its impact on pass-fail decision making* [Conference presentation]. National Council on Measurement in Education (NCME) Annual Meeting, San Francisco, CA, United States.
- Kamata, A., Turhan, A., & Darandari, E. (2003, April). *Estimating reliability for multidimensional composite scale scores* [Paper presentation]. American Educational Research Association Annual Meeting, Chicago, IL, United States.
- Kane, M., & Case, S. M. (2004). The reliability and validity of weighted composite scores. *Applied Measurement in Education*, 17, 221–240.
- Kenyon, D. M. (2006). *Development and field test of ACCESS for ELLs* [Technical Report No. 1]. WIDA, Center for Applied Linguistics, and the Board of Regents of the University of Wisconsin System.
- Kenyon, D. M., Ryu, J. R., & MacGregor, D. (2013). *Setting grade level cut scores for ACCESS for ELLs* [Technical report]. WIDA, Center for Applied Linguistics, and the Board of Regents of the University of Wisconsin System.

- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). Springer.
- Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement. *Journal of Educational Measurement*, 29, 285–307.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, 9(2), 202–220. <https://doi.org/10.1177/1094428105284919>
- Lee, W., Hanson, B. A., & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement*, 26, 412–432.
- Linacre, J. M. (1994). Sample size and item calibrations stability. *Rasch Measurement Transactions*, 7(4), 328.
- Linacre, J. M. (1999). Relating Cronbach and Rasch reliabilities. *Rasch Measurement Transactions*, 13(2), 696. <http://www.rasch.org/rmt/rmt132i.htm>
- Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85–106.
- Linacre, J. M. (2002b, Autumn). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878. <https://www.rasch.org/rmt/rmt162f.htm>
- Linacre, J. M. (2004). Optimizing rating scale category effectiveness. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 258–278). JAM Press.
- Linacre, J. M. (2006). Winsteps Rasch analysis (Version 3.60.1) [Computer software]. <https://www.winsteps.com>
- Linacre, J. M. (2020). *Reliability and separation of measures*. Winsteps. <https://www.winsteps.com/winman/reliability.htm>
- Linacre, J. M. (n.d.). *Displacement measures*. Winsteps. <https://www.winsteps.com/winman/displacement.htm>
- Livingston, S. A. (2018). *Test reliability—Basic concepts* [Research memorandum]. ETS Research Institute.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.

- Mantel, N., & Haenszel, W. (1959). Statistical aspect of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.
- Meyer, J. P. (2018). jMetrik [Computer software]. Retrieved from <https://itemanalysis.com/jmetrik-download/>
- Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, 17(4), 151–363.
- National Center on Educational Outcomes. (2021). *Universal design of assessments*. [https://nceo.info/Assessments/universal\\_design](https://nceo.info/Assessments/universal_design)
- Nunnally, J.C. (1978). An overview of psychological measurement. In: Wolman, B.B. (Ed.), *Clinical diagnosis of mental disorders* (pp. 97–146). Springer. [https://doi.org/10.1007/978-1-4684-2490-4\\_4](https://doi.org/10.1007/978-1-4684-2490-4_4)
- Price, L. R., Lurie, A., Raju, N., Wilkins, C., & Zhu, J. (2006). Conditional standard errors of measurement for composite scores on the Wechsler Preschool and Primary Scale of Intelligence – Third edition. *Psychological Reports*, 98(1), 237–252.
- Reise, S. P. (1999). Personality measurement issues viewed through the eyes of IRT. In S. E. Embretson, & S. L. Hershberger (Eds.). *The new rules of measurement: What every psychologist and educator should know* (pp. 219–240). Psychology Press.
- Rudner, L. (2001). Informed test component weighting. *Educational Measurement: Issues and Practice*, 20(1), 16–19.
- Sahakyan, N. (2020). *Generating alternate overall composite scale scores for English Learners with disabilities who are missing domain scores in the ACCESS for ELLs assessment* [Technical Report]. WIDA and the Board of Regents of the University of Wisconsin System.
- Sahakyan, N., & Poole, G. (2023, April). *Examining English learner testing, proficiency, and growth: Before, during, and “after” the COVID-19 pandemic* [Research report]. WIDA, Wisconsin Center for Educational Research, and the Board of Regents of the University of Wisconsin System. <https://wida.wisc.edu/resources/examining-english-learner-testing-proficiency-and-growth-before-during-and-after-covid-19>
- Stahl, J. A., & Muckle, T. (2007). Investigating drift displacement in Rasch item calibrations. *Rasch Measurement Transactions*, 21(3), 1126–1127.
- Thissen, D. (2000). Reliability and measurement precision. In H. Wainer, N. Dorans, D. Eignor, R. Flaugher, B. Green, R. Mislevy, L. Steinberg, & D. Thissen (Eds.), *Computerized adaptive testing: A primer* (2nd ed., pp. 159–184). Lawrence Erlbaum Associates.

- Turner, C.E. & Upshur, J.A. (2002). Rating scales derived from student samples: Effects of the scale maker and the student sample on scale content and student scores. *TESOL Quarterly*, 36(1), 49–70. <https://doi.org/10.2307/3588360>
- U.S. Department of Education. (2018). *A state's guide to the U.S. Department of Education's assessment peer review process*. Retrieved from <https://oese.ed.gov/files/2020/07/assessmentpeerreview.pdf>
- WIDA. (2007). *English language proficiency standards and resource guide, 2007 edition, prekindergarten through grade 12*. Board of Regents of the University of Wisconsin System.
- WIDA. (2012). *WIDA 2012 amplification of the English language development standards kindergarten–grade 12*. Board of Regents of the University of Wisconsin System.
- WIDA. (2020). *WIDA consortium English language proficiency assessment for grades 1–12 test and item design plan ACCESS for ELLs online annual summative and WIDA screener online*. Board of Regents of the University of Wisconsin System.
- WIDA. (2021a). *ACCESS for ELLs test administrator manual*. Board of Regents of the University of Wisconsin System.
- WIDA. (2021b). *ACCESS for ELLs district and school test coordinator manual*. Board of Regents of the University of Wisconsin System.
- WIDA. (2021c). *Test policy handbook for state education agencies*. Board of Regents of the University of Wisconsin System.
- Wright, B. D. & Douglas, G.A. (1975). *Best test design and self-tailored testing*. Research memorandum, Statistical Laboratory, Department of Education, University of Chicago.
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. MESA Press.
- Young, M. J., & Yoon, B. (1998). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment* (CSE Technical Report 475). Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education and Information Studies.
- Zieky, M. (1993). DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–347). Erlbaum.
- Zwick, R., & Bridgeman, B. (2014). Evaluating validity, fairness, and differential item functioning in multistage testing. In Y. Duanli, A. A. von Davier, & C. Lewis (Eds.), *Computer multistage testing: Theory and applications* (pp. 271–284). CRC Press.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233–251.

Zwick, R., Thayer, D. T., & Wingersky, M. (1993). *A simulation study of methods for addressing differential item functioning in computer-adaptive tests* (ETS Research Report RR-93-11). Educational Testing Service. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2333-8504.1993.tb01522.x>

## Acknowledgments

We would like to extend our appreciation to the many CAL and WIDA staff members who have supported this work, including the following:

**From CAL:** Tanya Bitterman, M.A.; Sofia Buitrago, MS.; Yage (Leah) Guo, Ph.D.; Victor Eyo Essien; Seong Eun Hong, Ph.D.; Michele Kawood, M.S.Ed.; Justin Kelly, Ph.D.; Reshmi Kumpakha, M.A.; Samantha Musser, M.A.; Aubrey Sahouria; Yu Lan Su, Ph.D.; Jasmine Tsai, M.Ed.; Yamei Wang, Ph.D.; Frank Wucinski, M.A.; Shu Jing Yen, Ph.D.

**From WIDA:** Kyoungwon Bishop, Ph.D.; Syed Abdul Hadi, M.Sc.