**HIGH-STAKES TESTING: SELECTED REFERENCES**

**(Last updated 25 February 2025)**

Ali, M. M., Hamid, M. O., & Hardy, I. (2020). Ritualisation of testing: Problematising high-stakes English-language testing in Bangladesh. *Compare: A Journal of Comparative and International Education*, *50*(4), 533-553.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, *36*(5), 258-267.

Au, W. (2016). Meritocracy 2.0: High-stakes, standardized testing as a racial project of neoliberal multiculturalism. *Educational Policy*, *30*(1), 39-62.

Au, W. (2022). *Unequal by design: High-stakes testing and the standardization of inequality*. Routledge.

Bach, A. J. (2020). High-stakes, standardized testing and emergent bilingual students in Texas: An overview of study findings and a call for action. *Texas Journal of Literacy Education*, *8*(1), 18-37.

Belzak, W. C. (2022). The multidimensionality of measurement bias in high‐stakes testing: Using machine learning to evaluate complex sources of differential item functioning. *Educational Measurement: Issues and Practice, 42*(1),1-10.

Booth, D. K. (2017). *The sociocultural activity of high stakes standardised language testing: TOEIC washback in a South Korean context*. Springer.

Booth, D. K., Booth, & Voogd. (2018). *Sociocultural activity of high stakes standardised language testing*. Springer.

Boori, A. A., Ghazanfari, M., Ghonsooly, B., & Baghaei, P. (2024). Cognitive diagnostic modeling analysis of the reading comprehension section of an Iranian high-stakes language proficiency test. *International Journal of Language Testing*, *14*(1), 17-33. DOI: [10.22034/ijlt.2023.399561.1256](https://doi.org/10.22034/ijlt.2023.399561.1256)

Chak, M. C. (2024). Washback in language learning strategies under high stakes language testing: A study of the Hong Kong secondary system. *rEFLections*, *31*(1), 1-24. https://doi.org/10.61508/refl.v31i1.269539

Cho, H. J., & Han, J. (2018). Teaching to the high-stakes testing in second language learning. *INTESOL Journal*, *15*(1), 49-65.

Coltrane, B. (2002). English language learners and high-stakes tests: An overview of the issues. ERIC Digest. ED470981 2002-11-00. <https://files.eric.ed.gov/fulltext/ED470981.pdf>

Coniam, D., & Falvey, P. (2007). High-stakes testing and assessment: English language teacher benchmarking. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 457-471). Springer.

Cizek, G. J. (2005). High-stakes testing: Contexts, characteristics, critiques, and consequences. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 23-54). Lawrence Erlbaum.

DelliCarpini, M., Ortiz-Marrero, F. W., & Sumaryono, K. (2010). Success with ELLs: ELLs at the center: Rethinking high-stakes testing. *The English Journal*, *99*(6), 93-96.

Deygers, B. (2017). Just testing: Applying theories of justice to high-stakes language tests. *ITL-International Journal of Applied Linguistics*, *168*(2), 143-163.

Dong, M. (2020). Structural relationship between learners’ perceptions of a test, learning practices, and learning outcomes: A study on the washback mechanism of a high-stakes test. *Studies in Educational Evaluation*, *64*, Article 100824. https://doi.org/10.1016/j.stueduc.2019.100824

Dong, M. (2022). Do school status and classroom English learning practices influence Chinese learners’ perceptions of a high-stakes English test? *Language Testing in Asia*, *12*(1), Article 28. <https://doi.org/10.1186/s40468-022-00182-5>

Fan, J. (2018). A survey of English language testing practices in China: Students’ and teachers’ Perspectives. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 283-300). Springer. <https://doi.org/10.1007/978-3-319-77177-9_16>

Fan, J., Frost, K., & Liu, B. (2020). Teachers’ involvement in high-stakes language assessment reforms: The case of Test for English Majors (TEM) in China. *Studies in Educational Evaluation*, *66*, Article 100898. https://doi.org/10.1016/j.stueduc.2020.100898

Farah, B. (2007). Students’ voices on their experiences with a high stakes language test. In S. Midraj, A. Jendli, & A. Sellami (Eds.) *Research in ELT contexts* (pp. 252-270).TESOL Arabia.

Feldman, M., Lazzara, E. H., Vanderbilt, A. A., & Diaz Granados, D. (2012). Rater training to support high‐stakes simulation‐based assessments. *Journal of Continuing Education in the Health Professions, 32*(4), 279-286. <https://doi.org/10.1002/chp.21156>

Fleming, K., & Shinjee, B. (2024). English high-stakes testing and constructing the ‘international’in Kazakhstan and Mongolia. *Applied Linguistics Review*, *15*(4), 1253-1275. <https://doi.org/10.1515/applirev-2022-0067>

French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, *88*(3), 893-918. https://doi.org/10.1007/s10734-023-01148-z

Frisén, L. B., Sundqvist, P., & Sandlund, E. (2021). Policy in practice: Teachers’ conceptualizations of L2 English oral proficiency as operationalized in high-stakes test assessment. *Languages, 6*(4), <https://doi.org/10.3390/languages6040204>.

Fung, D., & Chung, E. (2024). Defining language goals in EMI: Vocabulary demand in a high-stakes assessment in Hong Kong. *Language and Education*, *38*(2), 188-202. <https://doi.org/10.1080/09500782.2023.2219654>

Giouroukakis, V., & Honigsfeld, A. (2010). High‐stakes testing and English language learners: Using culturally and linguistically responsive literacy practices in the high school English classroom. *TESOL Journal*, *1*(4), 470-499.

Greene, J. P., Winters, M. A., & Forster, G. (2004). Testing high-stakes tests: Can we believe the results of accountability tests?. *Teachers College Record*, *106*(6), 1124-1144.

Gunzenhauser, M. G. (2003). High-stakes testing and the default philosophy of education. *Theory into Practice*, *42*(1), 51-58.

Haladyna, T. M., & Downing, S. M. (2004). Construct irrelevant variance in high stakes testing. *Educational Measurement: Issues and Practices, 23*(1), 17-27.

Hickman, L., Dunlop, P. D., & Wolf, J. L. (2024). The performance of large language models on quantitative and verbal ability tests: Initial evidence and implications for unproctored high‐stakes testing. *International Journal of Selection and Assessment*, *32*(4), 499-511. <https://doi.org/10.1111/ijsa.12479>

Hill, K., & McNamara, T. (2015). Validity inferences under high-stakes conditions: A response from language testing. *Measurement: Interdisciplinary Research & Perspectives*, *13*(1), 39-43.

Hilliard III, A. G. (2000). Excellence in education versus high-stakes standardized testing. *Journal of Teacher Education*, *51*(4), 293-304.

Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow?. *The Reading Teacher*, *54*(5), 482-492.

Honigsfeld, A., & Giouroukakis, V. (2011). High-stakes assessments and English language learners. *Delta Kappa Gamma Bulletin*, *77*(4), 6-10.

Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure?. *Theory into Practice*, *42*(1), 30-41.

Huang, J., Han, T., & Schnapp, K. (2012). Do high-stakes test really address English language learners’ learning needs? A discussion of issues, concerns, and implications. *International Journal of Learning and Development*, *2*(1), 499-508.

Huhta, A., Kalaja, P., & Pitkänen-Huhta, A. (2006). Discursive construction of a high-stakes test: The many faces of a test-taker. *Language Testing*, *23*(3), 326-350.

Isbell, D. R., & Kim, J. (2023). Developer involvement and COI disclosure in high-stakes English proficiency test validation research: A systematic review. *Research Methods in Applied Linguistics*, *2*(3). <https://doi.org/10.1016/j.rmal.2023.100060>

Jin, Y. (2011). Fundamental concerns in high-stakes language testing: The case of the college English test. *Journal of Pan-Pacific Association of Applied Linguistics*, *15*(2), 71-83.

Jones, B. D. (2007). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology*, *23*(2), 65-86.

Kearns, L. L. (2011). High-stakes standardized testing and marginalized youth: An examination of the impact on those who fail. *Canadian Journal of Education/Revue canadienne de l'éducation*, *34*(2), 112-130.

Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education*, *51*(4), 315-327.

Kremmel, B., Eberharter, K., Holzknecht, F., & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high-stakes test development. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 173-194). Springer.

LaFlair, G. T., & Staples, S. (2017). Using corpus linguistics to examine the extrapolation inference in the validity argument for a high-stakes speaking assessment. *Language Testing*, *34*(4), 451-475.

Lee, H., & Winke, P. (2013). The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing, 30*(1), 99-123.

Leo, A. (2023). Agency in constrained spaces: How English learners and their teachers negotiate the power of high‐stakes exams. *TESOL Quarterly*, *57*(2), 351-374. <https://doi.org/10.1002/tesq.3149>

Li, C., Kruger, L. J., Beneville, M., Kimble, E., & Krishnan, K. (2018). The unintended consequences of high-stakes testing on English-language learners: Implications for the practice of school psychology. *School Psychology Forum, 12*(3), 79-90. <https://eric.ed.gov/?id=EJ1195483>

Li, Q., & Chan, K. K. (2024). Test takers’ attitudes of using exam-oriented mobile application as a tool to adapt in a high-stakes speaking test. *Education and Information Technologies*, *29*(1), 219-237. https://doi.org/10.1007/s10639-023-12297-0

Liu, X., & Yu, J. (2021). Relationships between learning motivations and practices as influenced by a high-stakes language test: The mechanism of washback on learning. *Studies in Educational Evaluation*, *68*, 100967. <https://doi.org/10.1016/j.stueduc.2020.100967>

Luna, C., & Turner, C. L. (2001). The impact of the MCAS: Teachers talk about high-stakes testing. *The English Journal*, *91*(1), 79-87.

Madaus, G. F. (1988). The distortion of teaching and testing: High‐stakes testing and instruction. *Peabody Journal of Education*, *65*(3), 29-46.

Madaus, G., & Russell, M. (2010). Paradoxes of high-stakes testing. *Journal of Education*, *190*(1-2), 21-30.

Marchant, G. J. (2004). What is at stake with high stakes testing? A discussion of issues and research. *Ohio Journal of Science, 104*(2), 2-7. https://kb.osu.edu/server/api/core/bitstreams/77b6fa03-e461-5e19-8386-84933e88b909/content

Mei, Y., & Cheng, L. (2014). Scoring fairness in large-scale high-stakes English language testing: An examination of the National Matriculation English Test. In D. Conium (Ed.), English language education and assessment: Recent developments in Hong Kong and Chinese mainland (pp.171-187). Springer. doi: 10.1007/978-981-287-071-1\_11.

Mei, Y., & Cheng, L. (2019). Impact of a high-stakes test on education: Evidence from newspaper data. *Foreign Language Testing and Teaching* 《外语测试与教学》, 1, 1-15.

Morgan, H. (2016). Relying on high-stakes standardized tests to evaluate schools and teachers: A bad idea. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *89*(2), 67-72.

Moses, M. S., & Nanna, M. J. (2007). The testing culture and the persistence of high stakes testing reforms. *Education and Culture*, 55-72.

Neittaanmäki, R., & Lamprianou, I. (2024). Communal factors in rater severity and consistency over time in high-stakes oral assessment. *Language Testing*, *41*(3), 584-605. <https://doi.org/10.1177/02655322241239363>

Nishizawa, H. (2024). Authenticity of academic lecture passages in high-stakes tests: A temporal fluency perspective. *Language Testing*, *41*(4), 792-816. <https://doi.org/10.1177/02655322241262453>

Qi, L. (2004). Has a high-stakes test produced the intended changes. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171-190). Routledge.

Qi, L. (2005). Stakeholders’ conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, *22*(2), 142–173. https://doi.org/10.1191/0265532205lt300oa

Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high‐stakes English test in China. *Assessment in Education: Principles, Policy & Practice*, *14*(1), 51-74. <https://doi.org/10.1080/09695940701272856>

Riemenschneider, A., Weiss, Z., Schröter, P., & Meurers, D. (2024). The interplay of task characteristics, linguistic complexity, and language proficiency in high‐stakes English as a foreign language writing. *TESOL Quarterly*, *58*(2), 775-801. <https://doi.org/10.1002/tesq.3254>

Saadatara, A., Kiany, G., & Talebzadeh, H. (2023). Bundles to beat the band in high-stakes tests: Pedagogical applications of an exploratory investigation of lexical bundles across band scores of the IELTS writing component. *Journal of English for Academic Purposes*, *61*, 101208.

Schmidgall, J. E., & Powers, D. E. (2017). Technology and high‐stakes language testing. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 317-331). Wiley.

Shepard, L. A. (2002). The hazards of high-stakes testing. *Issues in Science and Technology*, *19*(2), 53-58.

Sinclair, J., Larson, E. J., & Rajendram, S. (2019). “Be a machine”: International graduate students’ narratives around high-stakes English tests. *Language Assessment Quarterly*, *16*(2), 236-252. https://doi.org/10.1080/15434303.2019.1628238

Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, *78*(2), 260-329.

Song, X., Cheng, L., & Klinger, D. (2015). DIF investigation across groups of gender and academic background in a large-scale high-stakes language test. Papers in Language Testing and Assessment, 4(1), 97-124.

Spann, P., & Kaufman, D. (2015). The negative effects of high-stakes testing. *Education Law and Policy*, *1*(1), 1-14.

Spring, R. (2025). Event conflation in high stakes testing: a comparison of usage and relationship to writing scores by language types. *Cognitive Linguistics*, (0). <https://doi.org/10.1515/cog-2024-0005>

Staples, S., Biber, D., & Reppen, R. (2018). Using corpus‐based register analysis to explore the authenticity of high‐stakes language exams: A register comparison of TOEFL iBT and disciplinary writing tasks. *The Modern Language Journal*, *102*(2), 310-332.

Tao, X., & Aryadoust, V. (2024). A multidimensional analysis of a high-stakes English listening test: A corpus-based approach. *Education Sciences*, *14*(2), 137-160. <https://doi.org/10.3390/educsci14020137>

Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education Today*, *26*(8), 662-671.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: can this be predicted or controlled?. *System*, *28*(4), 499-509.

West, G. B., & Thiruchelvam, B. (2024). “It's not their English”: Narratives contesting the validity of a high-stakes test. In M. R Salaberry, A. Weideman, & W. L. Hsu (Eds.), *Ethics and context in second language testing* (pp. 80-105). Routledge.

White, C. S., Sturtevant, E. G., & Dunlap, K. L. (2003). Perspective and beginning teachers’ perceptions of the influence of high stakes tests on their literacy-related instructional beliefs and decisions. *Reading Research and Instruction, 42*, 39-62.

Winke, P. (2011). Evaluating the validity of a high‐stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly*, *45*(4), 628-660.

Wright, W. E. (2002). The effects of high stakes testing in an inner-city elementary school: The curriculum, the teachers, and the English language learners. *Current Issues in Education*, *5*(5), 1-23.

Wright, W. E., & Choi, D. (2006). The impact of language and high-stakes testing policies on elementary school English language learners in Arizona. *Education Policy Analysis Archives*, *14*(13), 1-75.

Xerri, D., & Briffa, P. V. (Eds.). (2018). *Teacher involvement in high-stakes language testing*. Springer.

Xu, S. (2021). Processes and effects of test preparation for writing tasks in a high-stakes admission test in China: Implications for test takers. *Studies in Educational Evaluation*, *70*, Article 101015. https://doi.org/10.1016/j.stueduc.2021.101015

Yang, R. (2023). Is it possible to implement learning-oriented assessment principles in test preparation? Evidence from a high-stakes standardised EFL test in China. In S. W. Chong & H. Reinders (Eds.), *Innovation in learning-oriented language assessment* (pp. 27-43). Springer. <https://doi.org/10.1007/978-3-031-18950-0_3>

Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self theories. *Assessment in Education: Principles, Policy & Practice*, *21*(1), 71-89. https://doi.org/10.1080/0969594X.2012.757546

Zhan, Y., & Wan, Z. H. (2016). Test takers’ beliefs and experiences of a high-stakes computer-based English listening and speaking test. *RELC Journal*, *47*(3), 363-376.

Zhang, J. (2021). A moderated mediation analysis of the relationship between a high-stakes English test and test takers’ extracurricular English learning activities. *Language Testing in Asia*, *11*(1), Article 5. https://doi.org/10.1186/s40468-021-00120-x