

## DAFTAR PUSTKA

- Abduh, M. (2019). *Panduan Penulisan Soal HOTS-Higher Order Thinking Skills*. (Asrijanty & D. Hadiana, Eds.). Jakarta: Pusat Penilaian Pendidikan.
- Abosalem, Y. (2016). Assessment Techniques and Students' Higher-Order Thinking Skills. *International Journal of Secondary Education*, 4(1), 1. <https://doi.org/10.11648/j.ijsedu.20160401.11>
- Airasian, P. W., & Gullickson, A. (1994). Examination of Teacher Self-Assessment. *Journal of Personnel Evaluation in Education*, 8(2), 195–203.
- Al-absi, M. (2012). The Effect of Open-Ended Tasks – as An Assessment Tool- on Fourth Graders ' Mathematics Achievement , and Assessing Students ' Perspectives About it. *Jordan Journal of Educational Sciences*, 9(3), 345–351.
- Al-Samraie, H., Teo, T., & Abbas, M. (2013). Can Structured Representation Enhance Students' Thinking Skills For Better Understanding of E-Learning Content? *Computers and Education*, 69, 463–473. <https://doi.org/10.1016/j.compedu.2013.07.038>
- Al-zboon, H. S., Alnasraween, M. S., & Alkursheh, T. O. (2021). The Effect of the Percentage of Missing Data on Estimating the Standard Error of the Items' Parameters and the Test Information Function According to the Three-Parameter Logistic Model in the Item Response Theory. *İlköğretim Online*, 20(1). <https://doi.org/10.17051/ilkonline.2021.01.82>
- Alexander, B., Owen, S., & Thames, C. B. (2020). Exploring Differences and Relationships Between Online Formative and Summative Assessments in Mississippi Career and Technical Education. *Asian Association of Open Universities Journal*, 15(3), 335–349. <https://doi.org/10.1108/aaouj-06-2020-0037>
- Ali, A. &, & Zulkifli, Z. (2016). The Development Of Students' Capabilities in Higher Order Thinking Skill (HOTS) Through Science Education. *Jurnal Pemikir Pendidikan*, 7, 73–88. <https://doi.org/http://dx.doi.org/10.1093/neuonc/nox036>
- Ali, R. (2018). Rasch Model Validation Of Instrument To Measure Students Readiness To Embedded Systems Design Course, (February).
- Anderson. (2001). *Taxonomy for Learning, Teaching, and Assessing*:
- Andrade, H. G. (2000). Using Rubrics to Promote Thinking and Learning, 57(5), 1–7.
- Andrich, D. (1989). Distinctions Between Assumptions and Requirements in Measurement in the Social Sciences. *Mathematical and Theoretical Systems*, 7–16.
- Andrich, David. (1978a). A Rating Formulation for Ordered Response Categories. *Psychometrika*, 43(4), 561–573. <https://doi.org/10.1007/BF02293814>
- Andrich, David. (1978b). Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers. *Applied Psychological Measurement*, 2(4), 581–594. <https://doi.org/10.1177/014662167800200413>
- Andrich, David. (1988). A General Form of Rasch ' s Extended Logistic Model for Partial Credit Scoring. *Applied Measurement in Education*, 1(4), 279–297. <https://doi.org/10.1207/s15324818ame0104>
- Apino, E and Retnawati, H. (2017). Developing Instructional Design to Improve Mathematical Higher Order Thinking Skills of Students Developing Instructional Design to Improve Mathematical Higher Order Thinking Skills of Students, 8. <https://doi.org/10.1088/1742-6596/755/1/011001>
- Apino, E., & Retnawati, H. (2015). Developing Mathematical Higher Order Thinking Skills of Senior High School Students, 2011.
- Arhin, A. K. (2014). The Effect of Performance Assessment-Driven Instruction on the Attitude and Achievement of Senior High School Students in Mathematics in Cape Coast Metropolis , Ghana, 5(38), 109–117.
- ASLAN, C. (2017). Analysis of Attainments and Evaluation Questions in Sociology Curriculum according to the SOLO Taxonomy. *Eurasian Journal of Educational Research*, 69, 75–92.

- Atasoy, E., & Konyalıhatipoğlu, M. E. (2019). Investigation of Students' Holistic and Analytical Thinking Styles in Learning Environments assisted with Dynamic Geometry Software. *TeĖitim VBilim*, 44(199), 49–74. <https://doi.org/10.15390/eb.2019.8003>
- Azhar, K., Daud, M., Khidzir, N. Z., & Ismail, A. R. (2018). Validity and reliability of instrument to measure social media skills among small and medium entrepreneurs at Pengkalan Datu River. *International Journal of Development and Sustainability*, 7(3), 1026–1037.
- Baghaei, P., Yanagida, T., & Heene, M. (2017). Development Of A Descriptive Fit Statistic for the Rasch Model. *North American Journal of Psychology*, 19(1), 155–168.
- Barak, M., & Shakhman, L. (2008). Fostering Higher Order Thinking In Science Class: Teachers' Reflections. *Teachers and Teaching: Theory and Practice*, 14(3), 191–208. <https://doi.org/10.1080/13540600802006079>
- Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: a cross-disciplinary study. *Higher Education*, 74(1), 49–64. <https://doi.org/10.1007/s10734-016-0027-7>
- Becker, A. (2010). Examining Rubrics Used to Measure Writing Performance in US Intensive English Programs. *The CATESOL Journal*, 22(1), 113–130. Retrieved from <http://www.catesol.org/Becker113-130.pdf>
- Beesley, A. D., Clark, T. F., Dempsey, K., & Tweed, A. (2018). Enhancing Formative Assessment Practice and Encouraging Middle School Mathematics Engagement and Persistence. *School Science and Mathematics*, 118(1–2), 4–16. <https://doi.org/10.1111/ssm.12255>
- Bennedson, J., & Schrey-Niemenmaa, K. (2016). A Preliminary Case Study for Collaborative Quality Enhancement. *Turku University of Applied Sciences*, 173–185.
- Bernard, H. (2011). *Research Methods in Anthropology: Qualitative and Quantitative Approaches*.
- Bichi, A. A., Talib, R., Atan, N. A., & Ibrahim, H. (2019). Validation of a Developed University Placement Test Using Classical Test Theory and Rasch Measurement Approach. *International Journal of ADVANCED AND APPLIED SCIENCES*, 6(6), 22–29. <https://doi.org/10.21833/ijaas.2019.06.004>
- Biggs, J. B. (1992). Modes of Learning, Forms of Knowing, and Ways of Schooling.” In Neo-Piagetian Theories of Cognitive Development: Implications and Applications for Education. *Routledge*, 31–51.
- Biggs, J. (2011). Academic: SOLO taxonomy. Nettside: [http://www.johnbiggs.com.au/solo\\_taxonomy.html](http://www.johnbiggs.com.au/solo_taxonomy.html) [lesedato 1.
- Biggs, John B., & Collis, K. F. (1982). *Evaluation the Quality of Learning: the SOLO Taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.
- Biggs, John, & Collis, K. F. (1980). SOLO Taxonomy. *Education News*, 17(5), 19–23.
- Biggs, John, & Tang, C. (2011). Train-the-Trainers: Implementing Outcomes-Based Teaching and Learning in Malaysian Higher Education. *Malaysian Journal of Learning and Instruction*, 8, 1–19.
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., ... Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117–140. <https://doi.org/10.1177/1478210314566733>
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working Inside the Black Box: Assessment for Learning in the Classroom. *Phi Delta Kappan*, 86(1), 8–21. <https://doi.org/10.1177/003172170408600105>
- Black, P., & Wiliam, D. (2018). Classroom Assessment and Pedagogy. *Assessment in Education: Principles, Policy and Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Blake, N., Smeyers, P., Smith, R., & Standish, P. (2003). The Blackwell guide to the philosophy of education. *Choice Reviews Online*, 40(09), 40-5319-40–5319. <https://doi.org/10.5860/choice.40-5319>
- Bland, L. M., & Gareis, C. R. (2018). Performance Assessments: A Review of Defintions, Quality

Characteristics, and Outcomes Associated with Their Use in K-12 Schools. *Teacher Educator's Journal*, 52–69.

Bloom, B. S. (1956). *Taxonomy Of Education Objectives*.

Bolarinwa, O. (2015). Principles and Methods of Validity and Reliability Testing of Questionnaires Used in Social and Health Science Researches. *Nigerian Postgraduate Medical Journal*, 22(4), 195. <https://doi.org/10.4103/1117-1936.173959>

Boman, J., Currie, G., MacDonald, R., Miller-Young, J., Yeo, M., & Zettel, S. (2017). Overview of Decoding Across the Disciplines. *New Directions for Teaching and Learning*, 2017(150), 13–18. <https://doi.org/10.1002/tl.20234>

Bonafide, D. Y., Wulandari, R., Critical, P. L. T., & Rangi, N. L. (2018). Critical thinking skills profile of senior high school students in Biology learning. *International Conference on Science Education (ICoSEd)*, 6.

Bond, T., & Fox, C. M. (2015). *Applying the Rasch Model* (3rd Editio). New York: Routledge. <https://doi.org/https://doi.org/10.4324/9781315814698>

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates Publishers. Retrieved from <https://books.google.co.id/books?id=SLbBuAAZfZAC>

Bonk, C. J., & Smith, G. S. (1998). Alternative Instructional Strategies for Creative and Critical Thinking in the Accounting Curriculum. *Journal of Accounting Education*, 16(2), 261–293.

Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, 15(4), 1–7. <https://doi.org/10.1187/cbe.16-04-0148>

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch Analysis in the Human Sciences. *Rasch Analysis in the Human Sciences*, 1–482. <https://doi.org/10.1007/978-94-007-6857-4>

Borg, W. R., & Gall, M. D. (1984). Educational research: An introduction. *British Journal of Educational Studies*, 32(3).

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in Developmental Science: Situations, Shortcomings, Solutions, and Standards. *Developmental Review*, 33(4), 357–370. <https://doi.org/10.1016/j.dr.2013.08.003>

Bose, J., & Rengel, Z. (2009). A Model Formative Assessment Strategy to Promote Student-Centered Self-Regulated Learning in Higher Education. *US-China Educ Rev*, 6(12), 29–35. Retrieved from [http://eric.ed.gov/?id=ED511170%5Chttp://live.eprints.soton.ac.uk/72286/1/US-China\\_Education\\_Review\\_09-12.pdf#page=33](http://eric.ed.gov/?id=ED511170%5Chttp://live.eprints.soton.ac.uk/72286/1/US-China_Education_Review_09-12.pdf#page=33)

Brabrand, C., & Dahl, B. (2009). Using the SOLO Taxonomy to Analyze Competence Progression of University Science Curricula. *Higher Education*, 58(4), 531–549. <https://doi.org/10.1007/s10734-009-9210-4>

Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. Springer. London: # Springer Science+Business Media, LLC 2009. <https://doi.org/10.1016/b978-0-12-411355-8.50014-0>

Branch, R. M. (2013). Robert Maribe Branch, 24–27. <https://doi.org/10.13811/j.cnki.eer.2013.08.005>

Bransford, J. (1986). Teaching Thinking and Problem Solving. *American Psychologist*, 41(10), 1078–1089.

Bransford, J. D., Ann L. Brown, A., & Cocking, R. R. (2004). *How People Learn Brain, Mind, Experience, and School Committee*. *Psychology* (Vol. 116). Retrieved from <http://books.google.com/books?hl=en&lr=&id=U4XMcjDF9ZMC&oi=fnd&pg=PA8&dq=How+People+Learn&ots=FuL1p01kHD&sig=rnZ94Y5wiE3hHDZmxZeVyX3xu-k>

Brink, M., & Bartz, D. E. (2017). Effective Use of Formative Assessment by High School Teachers. *Practical Assessment, Research and Evaluation*, 22(8), 1–10.

- Brodin, E. M. (2016). Critical and Creative Thinking Nexus: Learning Experiences of Doctoral Students. *Studies in Higher Education*, 41(6), 971–989. <https://doi.org/10.1080/03075079.2014.943656>
- Brookhart, S M. (2007). *Assess HOTS In Your Classroom*. Virginia: ASCD Alexandria.
- Brookhart, Susan M. (2010). *How to Assess Higher Order Thinking Skills in your Classroom*. Virginia, USA: Alexandria, Virginia USA.
- Brookhart, Susan M. (2018). Appropriate Criteria: Key to Effective Rubrics, 3(April). <https://doi.org/10.3389/feduc.2018.00022>
- Brookhart, Susan M, & Nitko, A. J. (2011). Strategies for Constructing Assessments of Higher Order Thinking Skills. *Assessment of Higher Order Thinking Skills*, 327–359.
- Browne, M. N., & Freeman, K. (2000). Distinguishing Features of Critical Thinking Classrooms. *Teaching in Higher Education*, 5(3), 301–309. <https://doi.org/10.1080/713699143>
- Budi, K. (2018, April). Indonesia Kirim Guru Ke Korea Selatan untuk Pelajari HOTS. Retrieved from <https://edukasi.kompas.com/read/2018/04/23/08050091/indonesia-kirim-guru-ke-korea-untuk-pelajari-hots>
- Burnett, P. C. (1999). Assessing the Structure of Learning Outcomes from Counselling Using the SOLO Taxonomy: An Exploratory Study. *British Journal of Guidance and Counselling*, 27(4), 567–580. <https://doi.org/10.1080/03069889908256291>
- Bush, W. S., & Greer, A. S. (1999). *Mathematics Assessment: A Practical Handbook for Grades 9-12*. National Council of Teachers of Mathematics.
- Bygate, M. (2001). Effects of Task Repetition on the Structure and Control of Language.
- Bygate, M., & Samuda, V. (2009). Creating Pressure in Task Pedagogy: The Joint Roles of Field, Purpose, and Engagement Within the Interaction Approach. In *Multiple perspectives on interaction* (pp. 96–122). Routledge.
- Byrnes, J. P. (2008). Cognitive Development During Adolescence. *Blackwell Handbook of Adolescence*, (January), 227–246. <https://doi.org/10.1002/9780470756607.ch11>
- Caniglia, J., & Meadows, M. (2018). An Application of the Solo Taxonomy to Classify Strategies Used by Pre-Service Teachers to Solve “One Question Problems.” *Australian Journal of Teacher Education*, 43(9), 75–89. <https://doi.org/10.14221/ajte.2018v43n9.5>
- Cappelleri, J. C., Jason Lundy, J., & Hays, R. D. (2014). Overview of Classical Test Theory and Item Response Theory for the Quantitative Assessment of Items in Developing Patient-Reported Outcomes Measures. *Clinical Therapeutics*, 36(5), 648–662. <https://doi.org/10.1016/j.clinthera.2014.04.006>
- Carless, D. (2005). Prospects for The Implementation of Assessment for Learning. *Assessment in Education*, 12(1), 39–54. <https://doi.org/10.1080/0969594042000333904>
- Çelîk, D. (2009). Pre-Service Secondary School Teachers’ Use of Symbols and Algebraic Relationships in Turkey. In *Reforming Teaching and Learning* (pp. 165–179). Brill Sense.
- Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A Rasch Model Analysis on Secondary Students’ Statistical Reasoning Ability in Descriptive Statistics. *Procedia - Social and Behavioral Sciences*, 129, 133–139. <https://doi.org/10.1016/j.sbspro.2014.03.658>
- Chan, S. W., Looi, C. K., & Sumintono, B. (2020). Assessing computational Thinking Abilities Among Singapore Secondary Students: a Rasch Model Measurement Analysis. *Journal of Computers in Education*, (0123456789). <https://doi.org/10.1007/s40692-020-00177-2>
- Chappuis, S., & Stiggins, R. J. (2002). Classroom Assessment for Learning. *Educational Leadership*, 60, 40–43. <https://doi.org/10.1080/13530198908705494>
- Charles, C., Fulmer, G. W., Brown, G. T. L., & Tan, K. (2019). Value , Practice and Proficiency : Teachers’ Complex Relationship with Assessment for Learning. *Teaching and Teacher Education*, 80(January), 39–47. <https://doi.org/10.1016/j.tate.2018.12.022>

- Chen, I. H., Gamble, J. H., Lee, Z. H., & Fu, Q. L. (2020). Formative Assessment with Interactive Whiteboards: A One-Year Longitudinal Study of Primary Students' Mathematical Performance. *Computers and Education*, 150, 103833. <https://doi.org/10.1016/j.compedu.2020.103833>
- Chen, M.-H. (2016). Theoretical Framework for Integrating Higher-order Thinking into L2 Speaking. *Theory and Practice in Language Studies*, 6(2), 217. <https://doi.org/10.17507/tpls.0602.01>
- Chick, H. (1998). Cognition in the Formal Modes: Research Mathematics and the SOLO Taxonomy. *Mathematics Education Research Journal*, 10(2), 4–26. <https://doi.org/10.1007/BF03217340>
- Chinedu, C.C. & Kamin, Y., & S., O. O. (2015). Strategies For Improving Higher Order Thinking Skills In Teaching And Learning Of Design And Technology Education. *Journal of Technical Education and Training*, 7(2), 35–43.
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, 41(3), 178–194. <https://doi.org/10.1177/0146621616677520>
- Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The Perspectives of Chemists. *Science Education*, 93(1), 56–85. <https://doi.org/10.1002/sce.20292>
- Clarke, D. (1992). Activating Assessment Alternatives in Mathematics. *Arithmetic Teacher*. Retrieved from <http://eric.ed.gov/?id=EJ440152>
- Clements, M. D., & Cord, B. A. (2013). Assessment Guiding Learning: Developing Graduate Qualities in an Experiential Learning Programme. *Assessment and Evaluation in Higher Education*, 38(1), 114–124. <https://doi.org/10.1080/02602938.2011.609314>
- Cohen, A. (1994). Assessing Written Expression. *Assessing Language Ability in the Classroom*, 303–357.
- Collis, K. F., & Biggs, J. B. (1979). Classroom Examples of Cognitive Development Phenomena: The SOLO Taxonomy.
- Conn, C. A., Bohan, K. J., Pieper, S. L., & Musumeci, M. (2020). Validity Inquiry Process: Practical Guidance for Examining Performance Assessments and Building a Validity Argument. *Studies in Educational Evaluation*, 65(July 2018), 100843. <https://doi.org/10.1016/j.stueduc.2020.100843>
- Conrad, K. J., Conrad, K. M., Mazza, J., Riley, B. B., Funk, R., Stein, M. A., & Dennis, M. L. (2017). Dimensionality, Hierarchical Structure, Age Generalizability, and Criterion Validity of the GAIN's Behavioral Complexity Scale. *Physiology & Behavior*, 176(1), 139–148. <https://doi.org/10.1016/j.physbeh.2017.03.040>
- CoPo, A. R. I. (2016). Students Initial Knowledge State And Test Design: Towards A Valid And Reliable Test Instrument. *Journal of College Teaching & Learning (TLC)*, 12(3), 189. <https://doi.org/10.19030/tlc.v12i3.9310>
- Corcoran, C. A., Dershimer, E. L., & Tichenor, M. S. (2004). A Teacher's Guide to Alternative Assessment: Taking the First Steps. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 77(5), 213–218. <https://doi.org/10.3200/tchs.77.5.213-218>
- Cotton, K. (1991). Teaching Thinking Skills By Cotton, (November).
- Crane, P. K., Gibbons, L. E., Jolley, L., & Van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care*, 44(11 SUPPL. 3), 115–123. <https://doi.org/10.1097/01.mlr.0000245183.28384.ed>
- Creswell, J., & Clark, V. (2011). *Designing and Conducting Mixed Method Research. Analysing and interpreting data in mixed methods research.*
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* ERIC.
- Dagostino, L., Carifio, J., Bauer, J. D. C., Zhao, Q., & Hashim, N. H. (2015). Using Bloom's Revised Taxonomy to Analyze A Reading Comprehension Instrument. *Current Issues in Education*, 18(2).
- Dann, R. (2014). Assessment in Education : Principles , Policy & Practice Assessment as learning : blurring

- the boundaries of assessment and learning for theory , policy and practice. *Assessment in Education: Principles, Policy & Practice*, (April), 37–41. <https://doi.org/10.1080/0969594X.2014.898128>
- Darandari, E., & Murphy, A. (2013). Assessment of Student Learning (pp. 61–71). Riyadh, Saudi Arabia: Springer ScienceCBusiness Media Dordrecht. <https://doi.org/10.1007/978-94-007-6321-0>
- Darling-Hammond, L. (2017). Developing and Measuring Higher Order Skills: Models for State Performance Assessment Systems (Vol. 53, p. 54). <https://doi.org/10.1017/CBO9781107415324.004>
- Davidowitz, B., & Potgieter, M. (2016). Use of the Rasch Measurement Model to Explore the Relationship Between Content Knowledge and Topic-Specific Pedagogical Content Knowledge for Organic Chemistry. *International Journal of Science Education*, 38(9), 1483–1503. <https://doi.org/10.1080/09500693.2016.1196843>
- Dayal, H., Lingam, G., Sharma, L., Fitoo, B., & Sarai, V. (2018). Fijian Primary School leaders' Beliefs and Practices About Assessment. *Asia-Pacific Journal of Teacher Education*, 48(2), 168–186. <https://doi.org/10.1080/1359866X.2018.1548692>
- Desai, M. (2010). A Conceptual Framework for Rights-Based Preventative Approach for Psychosocial Well-Being in Childhood, 6, 47–69. [https://doi.org/10.1007/978-90-481-9066-9\\_3](https://doi.org/10.1007/978-90-481-9066-9_3)
- Devon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., ... Kostas-Polston, E. (2007). A Psychometric Toolbox For Testing Validity and Reliability. *Journal of Nursing Scholarship*, 39(2), 155–164. <https://doi.org/10.1111/j.1547-5069.2007.00161.x>
- Dikko, M. (2016). Establishing construct validity and reliability: Pilot testing of a qualitative interview for research in takaful (Islamic insurance). *Qualitative Report*, 21(3), 521–528.
- Division, C. of E. C. for C. C. E. C. M. L. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.
- Djaali, P.D., & Muljono, D. P. (2007). Pengukuran dalam Bidang Pendidikan. Jakarta.
- Donovan, M. S., & Donovan, M. S. (2013). in Education Systems, 317(April), 317–320. <https://doi.org/10.1126/science.1236180>
- Drost, E. A. (2004). Validity and Reliability in Social Science Research, 38(1), 105–125.
- Dubas, J. M., & Toledo, S. A. (2016). Taking higher order thinking seriously: Using Marzano's taxonomy in the economics classroom. *International Review of Economics Education*, 21, 12–20. <https://doi.org/10.1016/j.iree.2015.10.005>
- Dudley, D., & Baxter, D. (2009). Assessing Levels of Student Understanding in Pre-Service Teachers Using a Two-Cycle SOLO Model. *Asia-Pacific Journal of Teacher Education*, 37(3), 283–293. <https://doi.org/10.1080/13598660903052282>
- Earl, L. M. (2003). *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning* (Second). London: SAGE Publications Ltd. Retrieved from [https://books.google.co.id/books?hl=id&lr=&id=zYh2AwAAQBAJ&oi=fnd&pg=PP1&dq=Earl,+L.+\(2003\).+Assessment+as+Learning:+Using+Classroom+Assessment+to+Maximise+Student+Learning.+Thousand+Oaks,+CA,+Corwin+Press&ots=O7V5LhOAqN&sig=9u2uz-p9U-00oboxLFNsxmhw86w&re](https://books.google.co.id/books?hl=id&lr=&id=zYh2AwAAQBAJ&oi=fnd&pg=PP1&dq=Earl,+L.+(2003).+Assessment+as+Learning:+Using+Classroom+Assessment+to+Maximise+Student+Learning.+Thousand+Oaks,+CA,+Corwin+Press&ots=O7V5LhOAqN&sig=9u2uz-p9U-00oboxLFNsxmhw86w&re)
- Educaci, C. D. E. L. A., Use, T. H. E., Concept, O. F., For, M., Children, D., Skills, R., & Classroom, L. (n.d). Universidad de cuenca facultad de filosofía, letras y ciencias de la educación “The Use Of Concept Maps For Developing Children’s Reading and Writing Skills In A Foreign – Language Classroom.”
- Edwards, L. (2016). Education, Technology And Higher Order Thinking Skills. *AARE Conference*, 1–18.
- Eivind, K., & Ytterhaug, B. O. (2020). Measuring mathematical identity in lower secondary school. *International Journal of Educational Research*, 103(7491), 101620. <https://doi.org/10.1016/j.ijer.2020.101620>
- Ekawati, R., Junaedi, I., & Nugroho, S. E. (2013). Studi Respon Siswa Dalam Menyelesaikan Soal

Pemecahan Masalah Matematika Berdasarkan Taksonomi Solo. *Unnes Journal of Research Mathematics Education*, 2(2).

- Ekua, A. T. A.-S. (2016). Assessment, Student Learning and Classroom Practice: A Review. *Journal of Education and Practice*, 7(21), 1–6. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1109385&site=ehost-live>
- Ennis, R. H. (1989). Critical Thinking and Subject Specificity: Clarification and Needed Research. *Educational Researcher*, 18(3), 4–10. <https://doi.org/10.3102/0013189X018003004>
- Enu, J. (2021). Factors affecting teacher educators adoption of formative assessment strategies in the mathematics classroom. *Journal of Education and Learning (EduLearn)*, 15(4), 483–489. <https://doi.org/10.11591/edulearn.v15i4.20341>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2017). Comparison of Convenience Sampling and Purposive Sampling Comparison of Convenience Sampling and Purposive Sampling. (January 2016). <https://doi.org/10.11648/j.ajtas.20160501.11>
- Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The Effects of A Digital Formative Assessment Tool on Mathematics Achievement and Student Motivation. *Computers and Education*, 106, 83–96. <https://doi.org/10.1016/j.compedu.2016.12.001>
- Facione, P. A. (1990). Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations. *Eric*.
- Ferketich, S. (1990). Internal Consistency Estimates of Reliability. *Research in Nursing & Health*, 13(6), 437–440. <https://doi.org/10.1002/nur.4770130612>
- Fitrianawati, M. (2017). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik. *JPT: Jurnal Pendidikan Tematik*, 2(3), 316–322. Retrieved from <http://publikasiilmiah.ums.ac.id/handle/11617/9117>
- Forster, M. (2004). Higher Order. In *Duke Mathematical Journal* (Vol. 27, pp. 313–318). <https://doi.org/10.1215/s0012-7094-60-02728-9>
- Forza, C. (2002). Survey Research in Operations Management: A process-Based Perspective. *International Journal of Operations and Production Management*, 22(2), 152–194. <https://doi.org/10.1108/01443570210414310>
- Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The Effects on Self-Regulated learning, Self-Efficacy and Performance of Establishing Assessment Criteria With Students. *Studies in Educational Evaluation*, 53(March), 69–76. <https://doi.org/10.1016/j.stueduc.2017.03.003>
- Freitas, S., Prieto, G., Simões, M. R., & Santana, I. (2014). Psychometric properties of the Montreal Cognitive Assessment (MoCA): An analysis using the Rasch model, 1–37.
- Galbraith, P. (1993). Paradigms, Problems and Assessment: Some Ideological Implications, 73–74.
- Ghalib, T. K., & Al-Hattami, A. A. (2015). Holistic Versus Analytic Evaluation of EFL Writing: A Case Study. *English Language Teaching*, 8(7), 225–236. <https://doi.org/10.5539/elt.v8n7p225>
- Gibson, J. J. (2006). The Concept Of the Stimulus In Psychology. *American Psychologist*, 15(11), 694–703. <https://doi.org/10.1037/h0047037>
- Ginat, D., & Menashe, E. (2015). SOLO Taxonomy for Assessing Novices' Algorithmic Design. *SIGCSE 2015 - Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, 452–457. <https://doi.org/10.1145/2676723.2677311>
- Glatzeder, B. M., & Goel, V. (2010). *Towards a Theory of Thinking Building Blocks for a Conceptual Framework. Calcium in Living Cells*. <https://doi.org/http://dx.doi.org/10.1016/B978-0-12-374841-6.00015-3>
- Glazer, E. M. (2003). *Real-Life Math: Everyday Use Of Mathematical Concepts. Choice Reviews Online* (Vol. 40). <https://doi.org/10.5860/choice.40-4652>

- Goodrich Andrade, H. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education*, 4, 1–39.
- Goodwin, L. D., & Leech, N. L. (2003). The Meaning of Validity in the New Standards for Educational and Psychological Testing: Implications for Measurement Courses. *Measurement and Evaluation in Counseling and Development*, 36(3), 181–191. <https://doi.org/10.1080/07481756.2003.11909741>
- Govender, P. (2019). Formative Assessment as formative Pedagogy in Grade 3 Mathematics. *South African Journal of Childhood Education*, 9(1), 1–12. <https://doi.org/10.4102/sajce.v9i1.739>
- Green, A. (2018). Assessment of Learning and Assessment for Learning. *The TESOL Encyclopedia of English Language Teaching*, 1–6. <https://doi.org/10.1002/9781118784235.eelt0352>
- Grider, C. (1993). Foundations of Cognitive Theory: A Concise Review. *Information Analysis*, 15. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED372324>
- Guilleux, A., Blanchin, M., & Hardouin, J. (2014). Power and Sample Size Determination in the Rasch Model: Evaluation of the Robustness of a Numerical Method to Non-Normality of the Latent Trait. *Journal of Educational Measurement*, 9(1), 1–7. <https://doi.org/10.1371/journal.pone.0083652>
- Gustafson, K. L., & Branch, R. M. (2002). What is Instructional Design. *Trends and Issues in Instructional Design and Technology*, 2, 10–16.
- Guthrie, J. T. (2016). NIH Public Access, 28(2), 239–269. <https://doi.org/10.1007/s11145-014-9523-2.Comparing>
- Hadzhikoleva, S., Hadzhikolev, E., & Kasakliev, N. (2019). Using Peer Assessment To Enhance Higher Order Thinking Skills. *TEM Journal*, 8(1), 242–247. <https://doi.org/10.18421/TEM81-34>
- Hambleton, R. K., & Cook, L. L. (1977). Latent Trait Models and Their Use in the Analysis of Educational Test Data. *Journal of Educational Measurement*, 14(2), 75–96. Retrieved from <http://www.jstor.org/stable/1434009>
- Hambleton, R. K., Swaminathan, H., Rogers, D. J., & Hambleton, R. K. (1991). *Fundamentals of Item Response Theory Library of Congress Cataloging-in-Publication Data*.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
- Handayani, P., Murniati, & M, S. S. (2015). Analisis Argumentasi Peserta Didik Kelas X SMA Muhammadiyah 1 Palembang Dengan Menggunakan Model Argumentasi Toulmin. *Jurnal Inovasi Dan Pembelajaran Fisika*, 2(1), 60–68.
- Happs, J. C. (1985). Cognitive Learning Theory and Classroom Complexity. *Research in Science & Technological Education*, 3(2), 159–174. <https://doi.org/10.1080/0263514850030109a>
- Harlen, W. (2005). Teachers' Summative Practices and Assessment for Learning – Tensions and Synergies. *Curriculum Journal*, 16, 207–223. <https://doi.org/10.1080/09585170500136093>
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and Relationships Between Formative and Summative Assessment. *International Journal of Phytoremediation*, 21(1), 365–379. <https://doi.org/10.1080/0969594970040304>
- Harlen, W., & Winter, J. (2004). The Development of Assessment for Learning: Learning from the Case of Science and Mathematics. *Language Testing*, 21(3), 390–408. <https://doi.org/10.1191/0265532204lt289oa>
- Harsch, C., & Martin, G. (2013). Comparing Holistic and Analytic Scoring Methods: Issues of Validity and Reliability. *Assessment in Education: Principles, Policy and Practice*, 20(3), 281–307. <https://doi.org/10.1080/0969594X.2012.742422>
- Hartono, W., Hadi, S., Rosnawati, R., & Retnawati, H. (2022). Uji Kecocokan Model Parameter Logistik Soal Diagnosa Kemampuan Matematika Dasar. *JNPM (Jurnal Nasional Pendidikan Matematika)*, 6(1), 125. <https://doi.org/10.33603/jnpm.v6i1.5899>



- Hawk, T. F. (2009). *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Academy of Management Briarcliff Manor, NY.
- Hay, P. J., & Macdonald, D. (2008). (Mis) Appropriations of Criteria and Standards-Referenced Assessment in a Performance-Based Subject. *Assessment in Education: Principles, Policy and Practice*, 15(2), 153–168. <https://doi.org/10.1080/09695940802164184>
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, 18(3), 66–67. <https://doi.org/10.1136/eb-2015-102129>
- Hess, A. K. N., & Greer, K. (2016). Designing For Engagement: Using The ADDIE Model To Integrate High-Impact Practices Into An Online Information Literacy Course. *Communications in Information Literacy*, 10(2), 264–282.
- Hodaňová, J., & Nocar, D. (2016). Mathematics Importance In Our Life. *INTED2016 Proceedings*, 1(March), 3086–3092. <https://doi.org/10.21125/inted.2016.0172>
- Holmefur, M., Sundberg, K., Wettergren, L., & Langius-Eklöf, A. (2015). Measurement Properties of The 13-Item Sense of Coherence Scale Using Rasch Analysis. *Quality of Life Research*, 24(6), 1455–1463. <https://doi.org/10.1007/s11136-014-0866-6>
- Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of Two- and Three-Parameter Logistic Item Characteristic Curves: A Monte Carlo Study. *Applied Psychological Measurement*, 6(3), 249–260. <https://doi.org/10.1177/014662168200600301>
- Hussein, H. (2014). ASSESSING CONSTRUCT VALIDITY AND RELIABILITY OF COMPETITIVENESS SCALE USING RASCH MODEL APPROACH, 113–120.
- Irvine, J. (2017). A Comparison of Revised Bloom and Marzano's New Taxonomy of Learning. *Research in Higher Education Journal*, 33(08), 16. Retrieved from <http://www.aabri.com/copyright.html>
- Istiyono, E., Sunu, W., Dwandaru, B., & Megawati, I. (2018). Application of Bloomian and Marzanoian Higher Order Thinking Skills in the Physics Learning Assessment: an Inevitability. *International Conference on Learning Innovation*, 164(Icli 2017), 136–142.
- Jaganathan, P., & Subramaniam, I. (2016). Incorporating Higher Order Thinking Skills In Task-Based Learning For Malaysian Undergraduates. *International Journal of Contemporary Applied Sciences*, 3(2), 274–288. Retrieved from [www.ijcas.net](http://www.ijcas.net)
- Janssen, K. C., Phillipson, S., O'Connor, J., & Johns, M. W. (2017). Validation of the Epworth Sleepiness Scale for Children and Adolescents using Rasch analysis. *Sleep Medicine*, 33, 30–35. <https://doi.org/10.1016/j.sleep.2017.01.014>
- Jones, I., & Alcock, L. (2014). Peer Assessment Without Assessment criteria. *Studies in Higher Education*, 39(10), 1774–1787. <https://doi.org/10.1080/03075079.2013.821974>
- Jonsson, B., Norqvist, M., Liljekvist, Y., & Lithner, J. (2014). Learning Mathematics Through Algorithmic and Creative Reasoning. *Journal of Mathematical Behavior*, 36, 20–32. <https://doi.org/10.1016/j.jmathb.2014.08.003>
- Joshua, M. T. (2005). *Fundamentals of test and measurement in education*. Calabar: The.
- Kang, H. A., Su, Y. H., & Chang, H. H. (2018). A Note on Monotonicity of Item Response Functions for Ordered Polytomous Item Response Theory Models. *British Journal of Mathematical and Statistical Psychology*, 71(3), 523–535. <https://doi.org/10.1111/bmsp.12131>
- Kanjee, A., & Sayed, Y. (2013). Assessment Policy in Post-Apartheid South Africa: Challenges for Improving Education Quality and Learning. *Assessment in Education: Principles, Policy and Practice*, 20(4), 442–469. <https://doi.org/10.1080/0969594X.2013.838541>
- Karakoç, M. (2016). The Significance Of Critical Thinking Ability In Terms Of Education. *International Journal of Humanities and Social Science*, 6(7), 81–84. Retrieved from [http://www.ijhssnet.com/journals/Vol\\_6\\_No\\_7\\_July\\_2016/10.pdf](http://www.ijhssnet.com/journals/Vol_6_No_7_July_2016/10.pdf)
- Karhami, S. K. A., & Supriyati, Y. (2019). *Pengembangan KBM Kategori HOTS Pada Jenjang Pendidikan*

Dasar. (W. Sopandi, Ed.) (I). Bandung: Gapura Press.

- Kaspersen, E., Pepin, B., & Sikko, S. A. (2017). Measuring Student Teachers' Practices and Beliefs About Teaching Mathematics Using The Rasch Model. *International Journal of Research and Method in Education*, 40(4), 421–442. <https://doi.org/10.1080/1743727X.2016.1152468>
- Kastberg, D., Chan, J., Murray, G., & Gonzales, P. (2015). Performance of U.S. 15-Year-Old Students in Science, Reading, and Mathematics Literacy in an International Context. *National Center for Education Statistics*, 1–42.
- Kazerooni, E. A. (2001). Fundamentals of Clinical Research for Radiologists, (November), 993–999.
- Kemdikbud, P. (2021). Pembelajaran dan Asesmen, X–76.
- Kessler, A. M., Stein, M. K., & Schunn, C. D. (2015). Cognitive Demand of Model Tracing Tutor Tasks : Conceptualizing and Predicting How Deeply Students Engage. *Springer Science+Business Media Dordrecht*, 21. <https://doi.org/10.1007/s10758-015-9248-6>
- Kimberlin, C. I., & Winterstein, A. G. (2008). Validity and Reliability Of Measurement Instruments Used In Research, 65. <https://doi.org/10.2146/ajhp070364>
- Knoch, U. (2011). Rating Scales for Diagnostic Assessment of Writing: What Should they Look Like and Where Should the Criteria Come From? *Assessing Writing*, 16(2), 81–96. <https://doi.org/10.1016/j.asw.2011.02.003>
- Koller, I., Maier, M. J., & Hatzinger, R. (2015). An Empirical Power Analysis of Quasi-Exact Tests for the Rasch Model Measurement Invariance in Small Samples. <https://doi.org/10.1027/1614-2241/a000090>
- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *American Journal of Psychology*, 41(4), 219–225. <https://doi.org/10.1207/s15430421tip4104>
- Krulik, S., Rudnick, J. A., & Milou, E. (2003). *Teaching Mathematics in Middle School: A Practical Guide*. Allyn and Bacon.
- Kumar, R. (2012). *RESEARCH METHODOLOGY a step-by-step guide for beginners*. הגושע עלון (Vol. 66).
- Kunhertanti, K., & Heri Santosa, R. (2018). The Role of Problem Solving Approach in Building Students' Self Confidence on Learning Mathematics. *Advances in Social Science, Education and Humanities Research (ASSEHR)*, 160(Incomed 2017), 133–136. <https://doi.org/10.2991/incomed-17.2018.28>
- Kusuma, Merta Dhewa, E. a. (2017). The Development of Higher Order Thinking Skill (Hots) Instrument Assessment In Physics Study. *IOSR Journal of Research & Method in Education (IOSRJRME)*, 07(01), 26–32. <https://doi.org/10.9790/7388-0701052632>
- Lai, E. R. (2012). *Critical Thinking: A Literature Review*. *Transfusion* (Vol. 35). <https://doi.org/10.1046/j.1537-2995.1995.35395184278.x>
- Lambert, M. C., Garcia, A. G., January, S. A. A., & Epstein, M. H. (2018). The impact of English language learner status on screening for emotional and behavioral disorders: A differential item functioning (DIF) study. *Psychology in the Schools*, 55(3), 229–239. <https://doi.org/10.1002/pits.22103>
- Leenknecht, M., Wijnia, L., Köhler, M., Fryer, L., Rikers, R., Loyens, S., ... Rikers, R. (2020). Assessment & Evaluation in Higher Education Formative Assessment as Practice: the Role of Student's Motivation. *Assessment & Evaluation in Higher Education*, 0(0), 1–20. <https://doi.org/10.1080/02602938.2020.1765228>
- Leikin, R., Zazkis, R., & Meller, M. (2018). Research Mathematicians as Teacher Educators: Focusing on Mathematics for Secondary Mathematics Teachers. *Journal of Mathematics Teacher Education*, 21(5), 451–473. <https://doi.org/10.1007/s10857-017-9388-9>
- Lenderking, W. C. W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2013). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. <https://doi.org/10.1007/s11136-013-0487-5>
- Liaw, M. (2007). Content-Based Reading and Writing for Critical Thinking Skills in an EFL Context,

2(Summer), 45–87.

- Lim, H. L., & Noraini Idris. (2006). Assessing Algebraic Solving Ability of Form Four students. *International Electronic Journal of Mathematics Education*, 1(1), 55–76.
- Linacre, John M. (2002). Optimizing Rating Scale Category Effectiveness, 85–106.
- Linacre, John Michael. (1994). Sample Size and Item Calibration or Person Measure Stability. *Rasch Measurement Transactions*, 7(4), 328. Retrieved from <http://www.rasch.org/rmt/rmt74m.htm>
- Linarce, J. M. (2008). Transactions of the Rasch Measurement SIG The Rasch Model as a Construct Validation Tool. *Rasch Measurement Transaction*, 22(1), 1145–1162.
- Lister, R., Simon, B., Thompson, E., Whalley, J. L., & Prasad, C. (2006). Not Seeing the Forest for the Trees: Novice Programmers and the SOLO Taxonomy. *Working Group Reports on ITiCSE on Innovation and Technology in Computer Science Education 2006*, 118–122. <https://doi.org/10.1145/1140124.1140157>
- Liu, X., & Li, L. (2014). Assessment Training Effects on Student Assessment Skills and Task Performance in a Technology-Facilitated Peer Assessment. *Assessment and Evaluation in Higher Education*, 39(3), 275–292. <https://doi.org/10.1080/02602938.2013.823540>
- Lub, V. (2015). Validity in Qualitative Evaluation. *International Journal of Qualitative Methods*, 14(5), 160940691562140. <https://doi.org/10.1177/1609406915621406>
- M. Kovač, M., & Vickov, G. (2019). The Influence of Pre-Task Planning on Speed and Breakdown Fluency. *Journal for Foreign Languages*, 11(1), 87–102. <https://doi.org/10.4312/vestnik.11.87-102>
- Maciejewski, W., & Merchant, S. (2016). Mathematical Tasks , Study Approaches , and Course Grades in Undergraduate Mathematics : a Year-by- Year Analysis. *International Journal of Mathematical Education in Science and Technology*, 5211(March). <https://doi.org/10.1080/0020739X.2015.1072881>
- Maerten-rivera, J. L., Huggins-manley, A. C., Adamson, K., Lee, O., & Llosa, L. (2015). Development and Validation of a Measure of Elementary Teachers ' Science Content Knowledge in Two Multiyear Teacher Professional Development Intervention Projects. *JOURNAL OF RESEARCH IN SCIENCE TEACHING*, 52(3), 371–396. <https://doi.org/10.1002/tea.21198>
- Mahendra, I. W. E. (2016). Contextual Learning Approach and Performance Assessment in Mathematics Learning. *International Research Journal of Management, IT & Social Sciences*, 3(3), 7–15. <https://doi.org/10.21744/irjmis.v3i3.88>
- Maiorana, V. (1992). *Critical Thinking Across The Curriculum: Building the Analytical Classroom*. <https://doi.org/10.1360/zd-2013-43-6-1064>
- Mairing, J. P. (2014). E – 14 Students ' Concept Map s in Abstract Algebra. *International Seminar on Innovation in Mathematics and Mathematics Education 1st ISIM-MED*.
- Mansell, W., & James, M. (2009). the Assessment Reform Group. *Assessment in Schools. Fit for Purpose? A Commentary by the Teaching and Learning Research Programme*.
- Mardapi, D. (2012). Pengukuran Penilaian dan Evaluasi Pendidikan. *Yogyakarta: Nuha Medika*, 45.
- MARTON, F., & SÄLJÖ, R. (1976). On Qualitative Differences in Learning: I-Outcome and Process\*. *British Journal of Educational Psychology*, 46(1), 4–11. <https://doi.org/10.1111/j.2044-8279.1976.tb02980.x>
- Marzano, R. J., & Kendall, J. S. (2008). *Designing and Assessing Educational Objectives: Applying the New Taxonomy*. Corwin Press.
- Mascolo, M. F. (2008). The Concept of Domain in Developmental Analyses of Hierarchical Complexity. *World Futures*, 64(5–7), 330–347. <https://doi.org/10.1080/02604020802301170>
- Massof, R. W. (2012). Is The Partial Credit Model a Rasch Model? *Journal of Applied Measurement*. Massof, Robert W.: 6th Floor, 550 N. Broadway, Baltimore, MD, US, 21205, [rmassof@lions.med.jhu.edu](mailto:rmassof@lions.med.jhu.edu); Richard M Smith.

- Masters, G. N. (1982). A rasch model for partial Credit Scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Matsumura, L. C., Correnti, R., & Wang, E. (2015). Classroom Writing Tasks and Students' Analytic Text-Based Writing. *Reading Research Quarterly*, 50(4), 417–438. <https://doi.org/10.1002/rrq.110>
- McCallum, S., & Milner, M. M. (2020). The Effectiveness of Formative Assessment: Student Views and Staff Reflections. *Assessment and Evaluation in Higher Education*, 0(0), 1–16. <https://doi.org/10.1080/02602938.2020.1754761>
- Meerbaum-Salant, O., Armoni, M., & Ben-Ari, M. (Moti). (2013). Learning computer Science Concepts with Scratch. *Computer Science Education*, 23(3), 239–264. <https://doi.org/10.1080/08993408.2013.832022>
- Mertler, C. A. (2001). Designing Scoring Rubrics for Your Classroom, 7(2000), 2000–2001.
- Meyer, J. P., & Hailey, E. (2012). A study of Rasch, Partial Credit, and Rating Scale Model Parameter Recovery in WINSTEPS and jMetrik. *Journal of Applied Measurement*. Meyer, J. Patrick: Curry School of Education, P.O. Box 400265, 405 Emmet Street South, Charlottesville, VA, US, 22904, meyerjp@virginia.edu: Richard M Smith.
- Michael, A. A. (2017). Impact of Classroom Assessment , Management and Control on the Academic Performance of Secondary School Students in Akwa Ibom State. *Journal of Education and Practice*, 8(18), 30–36.
- Minbashian, A., Huon, G. F., & Bird, K. D. (2004). Approaches to Studying and Academic Performance in Short-Essay Exams. *Higher Education*, 47(2), 161–176. <https://doi.org/10.1023/B:HIGH.0000016443.43594.d1>
- Mishra, P., & Mehta, R. (2017). What We Educators Get Wrong About 21st-Century Learning: Results of a Survey. *Journal of Digital Learning in Teacher Education*, 33(1), 6–19. <https://doi.org/10.1080/21532974.2016.1242392>
- Mohajan, H. K. (2017). Two Criteria for Good Measurements in Research: Validity and Reliability. *Annals of Spiru Haret University. Economic Series*, 17(4), 59–82. <https://doi.org/10.26458/1746>
- Mohamad, M. M., Sulaiman, N. L., Sern, L. C., Mohd, K., & Salleh. (2015). Measuring the Validity and Reliability of Research Instruments. Retrieved from <https://pdf.sciencedirectassets.com/277811/1-s2.0-S1877042815X00413/1-s2.0-S1877042815047771/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEA8aCXVzL>
- Mohamad, N., Osman, K., Mastura, T., & Soh, T. (2011). Instrument Development for 21st Century Skills in Biology. *Procedia - Social and Behavioral Sciences*, 15, 1470–1474. <https://doi.org/10.1016/j.sbspro.2011.03.312>
- Moni, R. W., & Moni, K. B. (2008). Student Perceptions and Use of an Assessment Rubric for a Group Concept Map in Physiology. *American Journal of Physiology - Advances in Physiology Education*, 32(1), 47–54. <https://doi.org/10.1152/advan.00030.2007>
- Moskal, B. M. (2001). Scoring Rubrics : What , When and How ?, 7(2000), 2000–2001.
- Mosvold, R. (2006). Mathematics In Everyday Life A Study Of Beliefs And Actions. Retrieved from [http://scholar.google.no/scholar?as\\_q=&as\\_epq=&as\\_oq=&as\\_eq=&as\\_occt=any&as\\_sauthors=R+Mosvold&as\\_publication=&as\\_ylo=&as\\_yhi=&btnG=&hl=no&as\\_sdt=0%2C5#5](http://scholar.google.no/scholar?as_q=&as_epq=&as_oq=&as_eq=&as_occt=any&as_sauthors=R+Mosvold&as_publication=&as_ylo=&as_yhi=&btnG=&hl=no&as_sdt=0%2C5#5)
- Mulbar, U., Rahman, A., & Ahmar, A. S. (2017). Analysis of the Ability in Mathematical Problem-Solving Based on SOLO Taxonomy and Cognitive Style. *World Transactions on Engineering and Technology Education*, 15(1), 68–73. <https://doi.org/10.26858/wtetev15i1y2017p6873>
- Mulcahy, C. A., Krezmien, M. P., & Travers, J. (2016). Improving Mathematics Performance Among Secondary Students With EBD: A Methodological Review. *Remedial and Special Education*, 37(2), 113–128. <https://doi.org/10.1177/0741932515579275>
- Mulvia, R., Taufik Ramlan Ramalis, & Ridwan Efendi. (2021). Mendeteksi Keajegan Butir Tes Dengan Fungsi Informasi. *Jurnal Pendidikan Indonesia*, 2(1), 72–84. <https://doi.org/10.36418/japendi.v2i1.66>

- Mulyadi, Y., & Wikanengsih. (2022). Prototipe Mata Pelajaran Bahasa Indonesia, *11*(1), 47–60. Retrieved from file:///C:/Users/ASUS/Downloads/49632-75676647658-1-SM.pdf
- Muraki, E., & Bock, R. D. (1997). PARSCALE 3: IRT Based Test Scoring and Item Analysis for Graded Items and Rating Scales. *Chicago: Scientific Software.*
- Muraki, Eiji. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement, 16*(2), 159–176. <https://doi.org/10.1177/014662169201600206>
- Murphy, C., Bianchi, L., McCullagh, J., & Kerr, K. (2013). Scaling Up Higher Order Thinking Skills And Personal Capabilities In Primary Science: Theory-Into-Policy-Into-Practice. *Thinking Skills and Creativity, 10*, 173–188. <https://doi.org/10.1016/j.tsc.2013.06.005>
- Muruganatham, G. (2015). Developing of E-Content Package By Using ADDIE Model, *1*, 52–54.
- Musan, M. S. (2014). He Effect Of Dynamic Mathematics Learning Environment On The Solo Understanding Levels For Equations And Inequalities Of 8 Th Graders. *Mustafa Kemal University Journal of Graduate School of Social Sciences, 11*(26), 195–207. Retrieved from <https://dergipark.org.tr/en/download/article-file/183129>
- Nakakoji, Y., & Wilson, R. (2020). Interdisciplinary learning in mathematics and science: transfer of learning for 21st century problem solving at university. *Journal of Intelligence, 8*(3), 1–23. <https://doi.org/10.3390/jintelligence8030032>
- Nakano, T. D. C., & Primi, R. (2014). Rasch-Master’s Partial Credit Model in the Assessment of Children’s Creativity in Drawings. *Spanish Journal of Psychology, 17*(2), 1–16. <https://doi.org/10.1017/sjp.2014.36>
- Narayanan, S., & Adithan, M. (2015). Analysis Of Question Papers In Engineering Courses With Respect To Hots (Higher Order Thinking Skills). *American Journal of Engineering Education (AJEE), 6*(1), 1–10. <https://doi.org/10.19030/ajee.v6i1.9247>
- Nasab, F. G. (2015). Alternative versus Traditional Assessment, *2*(6), 165–178.
- National, T. H. E., & Press, A. (2000). *How People Learn ; Brain, Mind, Experience, and School : Expanded Edition.* (B. . . [et Al.], Ed.). Washington, D.C. <https://doi.org/10.17226/9853>
- Neal, L. (1999). Distance learning. *Conference on Human Factors in Computing Systems - Proceedings, 7*(3), 157–158. <https://doi.org/10.1145/632716.632810>
- Nehm, R. H., & Schonfeld, I. S. (2008). Measuring Knowledge Of Natural Selection: A Comparison Of The CINS, An Open-Response Instrument, And An Oral Interview. *Journal of Research in Science Teaching, 45*(10), 1131–1160. <https://doi.org/10.1002/tea.20251>
- Newton, B. G., & Martin, E. (2014). Blooming, SOLO taxonomy, and Phenomenography as Analytic Methods: Use In Higher Education and Relationships With Student Learning Approach, (March).
- Ng, S. W., Kwan, Y. W., & Huey Lei, K. H. (2020). Enhancing Student Learning and Teacher Development: Does “Assessment for Learning” Matter? *International Journal of Educational Reform, 29*(3), 275–292. <https://doi.org/10.1177/1056787920918595>
- Ní Fhloinn, E., & Carr, M. (2017). Formative Assessment in Mathematics for Engineering Students. *European Journal of Engineering Education, 42*(4), 458–470. <https://doi.org/10.1080/03043797.2017.1289500>
- Nitko, A. J. (1996). *Educational assessment of students* (Second). ERIC.
- Noble, H., & Smith, J. (2015). Issues Of Validity and Reliability In Qualitative Research. *Evidence-Based Nursing, 18*(2), 34–35. <https://doi.org/10.1136/eb-2015-102054>
- Norqvist, M. (2017). The Effect of Explanations on Mathematical Reasoning Tasks. *International Journal of Mathematical Education in Science and Technology, 52*11(August), 17. <https://doi.org/10.1080/0020739X.2017.1340679>
- Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: responding to issues

- regarding methodology, policy, and equity. *ZDM - Mathematics Education*, 50(4), 555–570. <https://doi.org/10.1007/s11858-018-0963-z>
- Núñez-Peña, M. I., Bono, R., & Suárez-Pellicioni, M. (2015). Feedback on Students' Performance: A possible Way of Reducing the Negative Effect of Math Anxiety in Higher Education. *International Journal of Educational Research*, 70, 80–87. <https://doi.org/10.1016/j.ijer.2015.02.005>
- Nunnally, J. C. (1978). *Psychometric Theory*. McGraw-Hill Book Company.
- O'Neill, G., & Murphy, F. (2010). Guide to Taxonomies of Learning. *UCD Teaching and Learning/Resources*. [Accessed ..., (January), 10. Retrieved from <http://www.ucd.ie/t4cms/UCDTLA0034.pdf>
- OECD. (2015). PISA 2015 Result in Focus. <https://doi.org/10.1596/28293>
- OECD. (2018). *PISA*.
- Okukawa, H. (2007). Vocational Certificate Curriculum in Thailand.pdf. Thailand. Retrieved from <https://search.proquest.com/openview/6adbd677e298a4df61ca4e4482592739/1?pq-origsite=gscholar&cbl=38579>
- Okyere, M., Kuranchie, A., Larbi, E., & Twene, C. (2018). *Essentials of assessment in schools*. Sunyani: Aduana Printing Press.
- Ölmez, İ. B., & Ölmez, S. B. (2019). Validation of the Math Anxiety Scale with the Rasch Measurement Model. *Mathematics Education Research Journal*, 31(1), 89–106. <https://doi.org/10.1007/s13394-018-0244-8>
- Olsen, L. W. (2003). *Essays on Georg Rasch and his*. University of Copenhagen. Retrieved from <https://www.rasch.org/olsen.pdf>
- Olsson, T. (2005). *Qualitative Assessment in Engineering Education*.
- Olusola, A., & Luneta, K. (2013). Assessing The Primary School Mathematics In South African : Literature And Teachers Perspective, 248–260.
- Othman, H., Ismail, N. A., Asshaari, I., Hamzah, F. M., & Nopiah, Z. M. (2015). Application of Rasch Measurement Model for Reliability Measurement Instrument in Vector Calculus Course. *Journal of Engineering Science and Technology*, 10(Spec. Issue 2 on UKM Teaching and Learning Congress 2013, May 2015), 77–83.
- Oyinloye, O. M., & Imenda, S. N. (2019). The Impact of Assessment for Learning on Learner Performance in Life Science. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(11). <https://doi.org/10.29333/ejmste/108689>
- Padmavathy, R. D., & Mareesh, K. (2013). Effectiveness of Problem Based Learning In Mathematics. *International Multidisciplinary E-Journal*, II(I), 45–51. Retrieved from [www.shreepublishing.com](http://www.shreepublishing.com)
- Palm, T. (2008). Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature. *Practical Assessment, Research and Evaluation*, 13(4), 1–11.
- Pandey, P., & Zimitat, C. (2007). Medical Students' Learning of Anatomy: Memorisation, Understanding and Visualisation. *Medical Education*, 41(1), 7–14. <https://doi.org/10.1111/j.1365-2929.2006.02643.x>
- Parsons, S. A., Malloy, A., Ward, A., & Cohen, S. (2015). Students' Engagement in Literacy Tasks, (December). <https://doi.org/10.1002/trtr.1378>
- Paul, B. R. W., & Elder, L. (2002). *Critical Thinking: Tools for Taking Charge of Your Professional and Personal Life*.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. ERIC. Retrieved from [https://scholar.google.co.id/scholar?hl=id&as\\_sdt=0%2C5&q=Pellegrino%2C+Chudowsky%2C+and+Glaser+%282001+assessment&oq=Pellegrino%2C+Chudowsky%2C+and+Glaser+%282001%29+as](https://scholar.google.co.id/scholar?hl=id&as_sdt=0%2C5&q=Pellegrino%2C+Chudowsky%2C+and+Glaser+%282001+assessment&oq=Pellegrino%2C+Chudowsky%2C+and+Glaser+%282001%29+as)

- Perlman, C. (2002). An Introduction To Performance Assessment Scoring Rubrics, in Understanding Scoring Rubrics. *College Park, MD: ERIC Clearinghouse on Assessment and Evaluation*, 5–13.
- Permana, R. H. (2019). Survei Kualitas Pendidikan PISA 2018: RI Sepuluh Besar dari Bawah, p. 1. Retrieved from <https://news.detik.com/berita/d-4808456/survei-kualitas-pendidikan-pisa-2018-ri-sepuluh-besar-dari-bawah>
- Peters, O., Körndle, H., & Narciss, S. (2018). Effects of a Formative Assessment Script on How Vocational Students Generate Formative Feedback to a Peer's or Their Own Performance. *European Journal of Psychology of Education*, 33(1), 117–143. <https://doi.org/10.1007/s10212-017-0344-y>
- Piaget, J. (1954). *The Construction of Reality in the Child*. New York, NY, US: Basic Books. <http://dx.doi.org/10.1037/11168-000>.
- Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2016). Implementation of Formative Assessment—Effects of Quality of Programme Delivery on Students' Mathematics Achievement and Interest. *Assessment in Education: Principles, Policy and Practice*, 25(2), 160–182. <https://doi.org/10.1080/0969594X.2016.1170665>
- Potter, M. K., & Kustra, E. (2012). A Primer on Learning Outcomes and the SOLO Taxonomy. *Course Design for Constructive Alignment*, (Winter 2012), 1–22.
- Pressley, M., & Allington, R. L. (2014). *Reading Instruction that Works: The Case for Balanced Teaching*. Guilford Publications.
- Prins, F. J., Kleijn, R. De, & Tartwijk, J. Van. (2015). Assessment & Evaluation in Higher Education Students' Use of a Rubric for Research Theses, 2938(November). <https://doi.org/10.1080/02602938.2015.1085954>
- Programme, T., Assessment, I. S., & Tables, I. (2018). Indonesia What 15-year-old Students in Indonesia Know and Can Do Figure 1. Snapshot of Performance in Reading , Mathematics and Science, 1–10.
- Purnomo, Y. W. (2014). Assessment-Based Learning : Sebuah Tinjauan untuk Meningkatkan Motivasi Belajar dan Pemahaman Matematis. *Sigma Journal*, VI(June).
- R.Lynn, M. (1986). Determination and Quantification of Content Validity. *Nursing Research*, 35, 382–386. <https://doi.org/10.1097/00006199-198611000-00017>
- R, A. L. (1985). Three Coefficients for Analysing Reliability and Validity of Rating. *Educational and Psychological Measurement*, 45, 131–142. <https://doi.org/10.1177/07399863870092005>
- Radmehr, F., & Drake, M. (2019). Revised Bloom's Taxonomy and Major Theories and Frameworks That Influence the Teaching, Learning, and Assessment of Mathematics: A Comparison. *International Journal of Mathematical Education in Science and Technology*, 50(6), 895–920. <https://doi.org/10.1080/0020739X.2018.1549336>
- Rahardja, U., Aini, Q., & Graha, Y. I. (2019). Validity of Test Instruments. *Journal of Physics: Conference Series*, 1364, 12. <https://doi.org/10.1088/1742-6596/1364/1/012050>
- Rahayu, W., Putra, M. D. K., Iriyadi, D., Rahmawati, Y., & Koul, R. B. (2020). A Rasch and factor analysis of an Indonesian version of the Student Perception of Opportunity Competence Development (SPOCD) questionnaire. *Cogent Education*, 7(1). <https://doi.org/10.1080/2331186X.2020.1721633>
- Rahman, A. (2018). PEMBELAJARAN MATEMATIKA UNTUK MENINGKATKAN HIGHER ORDER THINKING SKILLS ( HOTS ). *Seminar Nasional Pendidikan Matematika (SENPIKA)*, (1), 1–11.
- Raj Acharya, B. (2017). Factors Affecting Difficulties in Learning Mathematics by Mathematics Learners. *International Journal of Elementary Education*, 6(2), 8. <https://doi.org/10.11648/j.ijeeedu.20170602.11>
- Rasch, G. (1960). Probabilistic Models For Some Intelligence And Attainment Tests. Retrieved from <https://raschmeas.info/rasch-references-basics-for-any-library-or-beach/>
- Reckase, M. D. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational Statistics*, 4(3), 207–230. <https://doi.org/10.3102/10769986004003207>

- Reddy, Y. M., & Andrade, H. (2010). Assessment & Evaluation in Higher Education A Review of Rubric Use in Higher Education. *Assessment & Evaluation in Higher Education*, 35(May 2009), 37–41. <https://doi.org/10.1080/02602930902862859>
- Reeve, B. B., & Fayers, P. (2005). Applying Item Response Theory Modelling for Evaluating Questionnaire Item and Scale Properties. *Assessing Quality of Life in Clinical Trial: Methods and Practice*, 55–74. <https://doi.org/10.1007/s11136-007-9198-0>
- Reise, S. P. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG, 27(2), 133–144.
- Resnick, L. B. (1987). *Education and learning to think*. National Academy Press.
- Resnick, Lauren B., & Resnick, D. P. (1992). Assessing the Thinking Curriculum: New Tools for Educational Reform. *Changing Assessments*, 37–75. [https://doi.org/10.1007/978-94-011-2968-8\\_3](https://doi.org/10.1007/978-94-011-2968-8_3)
- Retnawati, H. (2013). Pendeteksian Keberfungsian Butir Pembeda Dengan Indeks Volume Sederhana Berdasarkan Teori Respons Butir Multidimensi. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 17(2), 275–286. <https://doi.org/10.21831/pep.v17i2.1700>
- Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya* (1st ed.). Yogyakarta: Nuha Medika.
- Retnawati, H. (2017). *Validitas Reliabilitas dan Karakteristik Butir* (2nd ed.). Yogyakarta: Parama Publishing.
- Retnawati, P. and. (2018). Urgency of Higher Order Thinking Skills ( HOTS ) Content Analysis in Mathematics Textbook Urgency of Higher Order Thinking Skills ( HOTS ) Content Analysis in Mathematics Textbook.
- Richey, R. C. (n.d.). Klein.(2007). Design and Development Research. London: Lawrence Erlbaum Associates. Inc.
- Rooney, C. (2012). How Am I Using Inquiry-Based Learning to Improve My Practice and to Encourage Higher Order Thinking Among My Students of Mathematics ? *Educational Journal of Living Theories*, 5(2), 99–127.
- Runnels, J. (2012). Using The Rasch Model To Validate A Multiple Choice English Achievement Test. *International Journal of Language Studies*, 6(4), 141–153.
- Rusman, E., & Dirks, K. (2017). Developing rubrics to assess complex (generic) skills in the classroom: How to distinguish skills' mastery levels? *Practical Assessment, Research and Evaluation*, 22(12), 1–9.
- Rust, C. (2002). The Impact of Assessment on Student Learning: How Can the Research Literature Practically Help to Inform the Development of Departmental Assessment Strategies and Learner-Centred Assessment Practices? *Active Learning in Higher Education*, 3(2), 145–158. <https://doi.org/10.1177/1469787402003002004>
- Ruthven, K. (2011). Conceptualising Mathematical Knowledge in Teaching. In *Mathematical knowledge in teaching* (pp. 83–96). Springer.
- Şahin, M. (2019). Classroom Response Systems as a Formative Assessment Tool: Investigation into Students' Perceived Usefulness and Behavioural Intention. *International Journal of Assessment Tools in Education*, 6(4), 693–705. <https://doi.org/10.21449/ijate.576249>
- Said, R. F. M. (2016). *Application of Rasch Measurement Model in Evaluating Student Performance for Foundation of Computing II. 7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings*. [https://doi.org/10.1007/978-981-287-664-5\\_58](https://doi.org/10.1007/978-981-287-664-5_58)
- Saidi, S. S., & Siew, N. M. (2019). Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument using the Rasch Measurement Model. *International Electronic Journal of Mathematics Education*, 14(3), 535–546. <https://doi.org/10.29333/iejme/5755>
- Sajidan, & Afandi. (2017). Pengembangan Model Pembelajaran IPA Untuk Memberdayakan Keterampilan Berpikir Tingkat Tinggi. *Prosiding Seminar Nasional Pendidikan Sains (SNPS)*, 21, 15–27.
- Samritin; Suryanto. (2016). Developing An Assessment Instrument Of Junior High School Students' Higher



Order Thinking Skills In Mathematics, 2(1), 71–78.

- Sang, G., Liang, J. C., Chai, C. S., Dong, Y., & Tsai, C. C. (2018). Teachers' actual and preferred perceptions of twenty-first century learning competencies: a Chinese perspective. *Asia Pacific Education Review*, 19(3), 307–317. <https://doi.org/10.1007/s12564-018-9522-0>
- Santun Naga, D. (2012). *Teori Sekor Pada Pengukuran Mental* (2nd ed.). Jakarta: PT Nagarani Citrayasa.
- Scaltritti, M., Dufau, S., & Grainger, J. (2018). Stimulus Orientation and the First-Letter Advantage. *Acta Psychologica*, 183(September 2017), 37–42. <https://doi.org/10.1016/j.actpsy.2017.12.009>
- Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71(February), 101094. <https://doi.org/10.1016/j.stueduc.2021.101094>
- Scully, D. (2017). Constructing Multiple-Choice Items to Measure Higher-Order Thinking. Practical Assessment, Research & Evaluation. *Practical Assessment, Research & Evaluation*, 22(4). Retrieved from <http://pareonline.net/getvn.asp?v=22&n=4>
- Selwyn, N. (2007). The Use Of Computer Technology In University Teaching and Learning: A critical Perspective. *Journal of Computer Assisted Learning*, 23(2), 83–94. <https://doi.org/10.1111/j.1365-2729.2006.00204.x>
- Serow, P., Callingham, R., & Tout, D. (2016). Assessment of Mathematics Learning: What Are We Doing? *Research in Mathematics Education in Australasia 2012-2015*, 1–378. <https://doi.org/10.1007/978-981-10-1419-2>
- Sharifah, R., & Faaizah, S. (2015). The Development of Online Project Based Collaborative Learning using ADDIE Model. *Procedia - Social and Behavioral Sciences*, 195, 1803–1812. <https://doi.org/10.1016/j.sbspro.2015.06.392>
- Shavelson, R. J., Domingue, B. W., Mariño, J. P., Molina Mantilla, A., Morales Forero, A., & Wiley, E. E. (2016). On the Practices and Challenges of Measuring Higher Education Value Added: the Case of Colombia. *Assessment and Evaluation in Higher Education*, 41(5), 695–720. <https://doi.org/10.1080/02602938.2016.1168772>
- Shavelson, R. J., Zlatkin-Troitschanskaia, O., & Mariño, J. P. (2018). International Performance Assessment of Learning in Higher Education (iPAL): Research and Development, 193–214. [https://doi.org/10.1007/978-3-319-74338-7\\_10](https://doi.org/10.1007/978-3-319-74338-7_10)
- Shukla, D., & Dungsungnoen, P. (2016). Students Perceived Level and Teachers Teaching Strategies of Higher Order Thinking Skills; A Study on Higher Educational Institutions in Thailand. *Journal of Education and Practkice*, 7(12), 211–219.
- Siagian, S. (2018). The Development of Assessment Instrument for the Character Education at the State University of Medan. *Advances in Social Science, Education and Humanities Research*, 200(3rd Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2018) The), 889–895.
- Sidenvall, J., Lithner, J., & Jäder, J. (2015). Students' Reasoning in Mathematics Textbook Task-Solving. *International Journal of Mathematical Education in Science and Technology*, 46(4), 533–552. <https://doi.org/10.1080/0020739X.2014.992986>
- Silber-Varod, V., Eshet-Alkalai, Y., & Geri, N. (2019). Tracing research trends of 21st-century learning skills. *British Journal of Educational Technology*, 50(6), 3099–3118. <https://doi.org/10.1111/bjet.12753>
- Simamora, K. A. L. G., & Tilaar, A. L. F. (2021). Analisis Kemampuan Literasi Matematika Ditinjau Dari Penggunaan Soal-Soal Matematika Tipe HOTS. *MARISEKOLA: Jurnal Matematika Riset Edukasi Dan Kolaborasi*, 2(1), 23–30. <https://doi.org/10.53682/marisekola.v2i1.1139>
- Smit, R., & Birri, T. (2014). Assuring the Quality of Standards-Oriented Classroom Assessment With Rubrics for Complex Competencies. *Studies in Educational Evaluation*, 43(March), 5–13. <https://doi.org/10.1016/j.stueduc.2014.02.002>

- Smith, T. W., & Colby, S. A. (2007). Teaching for Deep Learning. *The Clearing House : A Journal of Educational Strategies*, *Teaching for Deep Learning*, (May 2007), 1–8. <https://doi.org/10.3200/TCHS.80.5.205-210>
- Sondergeld, T. A., & Johnson, C. C. (2014). Using Rasch Measurement for the Development and Use of Affective Assessments in Science Education Research. *Science Education*, *98*(4), 581–613. <https://doi.org/10.1002/sce.21118>
- Srirahayu, R. R. Y., & Arty, I. S. (2018). Pengembangan Instrumen Experiment Performance Assessment Untuk Menilai Keterampilan Proses Sains Dan Kerja Sama, *22*(2), 168–181.
- Stålné, K., Kjellström, S., & Utriainen, J. (2016). Assessing Complexity in Learning Outcomes – a Comparison Between the SOLO Taxonomy and the Model of Hierarchical Complexity. *Assessment and Evaluation in Higher Education*, *41*(7), 1033–1048. <https://doi.org/10.1080/02602938.2015.1047319>
- Star, J. R., & Stylianides, G. J. (2013). Procedural and Conceptual Knowledge: Exploring the Gap Between Knowledge Type and Knowledge Quality. *Canadian Journal of Science, Mathematics and Technology Education*, *13*(2), 169–181. <https://doi.org/10.1080/14926156.2013.784828>
- Stein, M. K., & Lane, S. (1996). Instructional Tasks and The Development of Student Capacity to Think and Reason: An Analysis of The Relationship Between Teaching and Learning in a Reform Mathematics Project. *International Journal of Phytoremediation*, *21*(1), 50–80. <https://doi.org/10.1080/1380361960020103>
- Stein, M. K., & Smith, M. S. (1998). Mathematical Tasks as a Framework for Reflection : From Research to Practice. *Mathematics Teaching in the Middle School*, *3*(4), 268–275.
- Stenner, A. J., Fisher, W. P., Stone, M. H., & Burdick, D. S. (2013). Causal Rasch models. *Frontiers in Psychology*, *4*(AUG), 1–14. <https://doi.org/10.3389/fpsyg.2013.00536>
- Stevens, D. D., Levi, A., & State, P. (2006). Essays on Teaching Excellence Leveling the Field: Using Rubrics to Achieve Greater Equity in Teaching and Grading. *Development*, *17*(1), 2005–2006.
- Stiggins, R. J. (2010). Assessment Crisis: The Absence Of Assessment FOR Learning. *Special Section On Assessment*, 19–28. Retrieved from <http://beta.edtechpolicy.org/CourseInfo/edhd485/AssessmentCrisis.pdf>
- Stone, C. A., & Zhang, B. (2003). Assessing Goodness of Fit of Item Response Theory Models: A Comparison of Traditional and Alternative Procedures. *Journal of Educational Measurement*, *40*(4), 331–352. <https://doi.org/10.1111/j.1745-3984.2003.tb01150.x>
- Sudaryono. (2012). the Various Methods of Detecting the Existence of Differential Item Function (Dif) Item National Exam With Classical Test Theory. *Jurnal Pendidikan Dan Kebudayaan*, *18*(2), 132–144.
- Sullivan, P., Askew, M., Cheeseman, J., & Clarke, D. (2014). Supporting Teachers in Structuring Mathematics Lessons Involving Challenging Tasks, (April 2015). <https://doi.org/10.1007/s10857-014-9279-2>
- Sumintoro, Bambang, Widhiarso, W. (2013). *Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. (B. Sumintoro, Ed.) (II). Cimahi.
- Sumintoro, Bambang, Widhiarso, W. (2015). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. (B. Trim, Ed.) (II). Cimahi.
- Supratiknya. (2006). *Penilaian Hasil Belajar dengan Teknik Nontes*. *European Heart Journal* (Vol. 28).
- Susetyo, B. (2015). *Prosedur Penyusunan dan Analisis Tes*. Bandung: PT Refika Aditama. Retrieved from <https://refika.co.id/pendidikan/206-prosedur-penyusunan-analisis-tes.html>
- Swartz, Robert J.; Parks, S. (1994). *Infusing the Teaching of Critical and Creative Thinking into Content Instruction: A Lesson Design Handbook for the Elementary Grades*. ERIC Number: ED407059. Retrieved from <https://eric.ed.gov/?id=ED407059>
- Swartz, R., & Mcguinness, C. (2014). Developing and Assessing Thinking Skills. *The International*

*Baccalaureate Project Final Report Part 2 Evaluation of Current IB Programmes*, (February), 1–68.

- Swerdlik, M. E., & Cohen, R. J. (2005). *Psychological testing and assessment: An introduction to tests and measurement*. Boston: McGraw-Hill.
- Tabatabaee-Yazdi, M. et. al. (2018). Development and Validation of a Teacher Success Questionnaire Using the Rasch Model. *International Journal of Instruction*, 11(2), 129–144. <https://doi.org/10.12973/iji.2018.11210a>
- Taherdoost, H. (2018). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*, (January 2016). <https://doi.org/10.2139/ssrn.3205040>
- Tan, S. Y., & Siti Hajar, H. (2015). Effective Teaching of Higher-Order Thinking (HOT) in Education. *The Online Journal of Distance Education and E-Learning (TOJDEL)*, 3(2), 41–47.
- Tara, M. (2005). Assessment - Summative and Formative- Some Theoretical Reflections. *British Journal of Educational Studies*, 53(4), 466–478.
- Tejeda, S., & Gallardo, K. (2017). Performance Assessment on High School Advanced Algebra, 12(9), 777–798.
- Teo, T. W., & Goh, W. P. J. (2019). Assessing lower track students' learning in science inference skills in Singapore. *Asia-Pacific Science Education*, 5(1). <https://doi.org/10.1186/s41029-019-0033-z>
- Timofte, R. S., & Siminiciuc, L. (2018). Utilisation of Rasch Model for The Analysis of an Instrument Developed by Mapping Items to Cognitive Levels of Marzano Taxonomy. *Acta Didactica Napocensia*, 11(2), 71–78. <https://doi.org/10.24193/adn.11.2.6.72>
- Tomas, C., Whitt, E., Lavelle-Hill, R., & Severn, K. (2019). Modeling Holistic Marks With Analytic Rubrics. *Frontiers in Education*, 4(September), 1–19. <https://doi.org/10.3389/educ.2019.00089>
- Tornabene, R. E., Lavington, E., & Nehm, R. H. (2018). Testing Validity Inferences for Genetic Drift Inventory Scores Using Rasch Modeling and Item Order Analyses. *Evolution: Education and Outreach*, 11(1), 6. <https://doi.org/10.1186/s12052-018-0082-x>
- Torrance, H. (2007). Assessment as Learning? How The Use of Explicit Learning Objectives, Assessment Criteria and Feedback in Post-Secondary Education and Training Can Come to Dominate Learning. *Assessment in Education: Principles, Policy and Practice*, 14(3), 281–294. <https://doi.org/10.1080/09695940701591867>
- Trust, T., & Pektas, E. (2018). Using the ADDIE Model and Universal Design for Learning Principles to Develop an Open Online Course for Teacher Professional Development. *Journal of Digital Learning in Teacher Education*, 34(4), 219–233. <https://doi.org/10.1080/21532974.2018.1494521>
- Ülger, K. (2016). The Relationship between Creative Thinking and Critical Thinking Skills of Students Öğrencilerin Yaratıcı Düşünme ve Eleştirel Düşünme Becerileri Arasındaki İlişki, 31(4), 695–710. <https://doi.org/10.16986/HUJE.2016018493>
- Uli Sihombing, Rosita; S. Naga, Dali; Rahayu, W. (2019). A Rasch Model Measurement Analysis On Science Literacy Test Of Indonesian Student : Smart Way To Improve Learning Assessment., 6(1), 44–55.
- Ulumudin, I., & Fujianita, S. (2019). The Implementation of Attitude Assessment in Curriculum 2013 at Elementary Schools. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 23(1), 46–56. <https://doi.org/10.21831/pep.v23i1.23391>
- Umar, H. (2021). Analisis Karakteristik Stimulus dan Level Kognitif Soal Um-Ptkin Tahun 2019. *Jurnal Pendidikan Matematika Universitas Lampung*, 9(2), 184–196. <https://doi.org/10.23960/mtk/v9i2.pp184-196>
- Urbina, A. A. dan S. (2007). *Tes Psikologi*. Jakarta: PT Indeks.
- Vallori, A. B. (2014). Meaningful Learning In Practice, 3(4), 199–209. <https://doi.org/10.15640/jehd.v3n4a18>

- Van Zile-Tamsen, C. (2017). Using Rasch Analysis to Inform Rating Scale Development. *Research in Higher Education*, 58(8), 922–933. <https://doi.org/10.1007/s11162-017-9448-0>
- VanderStoep, Scott W; Johnston, D. D. (n.d.). *Research Methods For Everyday Life*.
- Verhelst, N. D., & Verstralen, H. H. F. M. (2008). Some Considerations on The Partial Credit Model. *Psicologica*, 29(2), 229–254.
- Wahyuningsih, S. (2021). Using the Rasch's Partial Credit Model to Analyze the Quality of an Essay Math Test. *Proceedings of the 1st International Conference on Mathematics and Mathematics Education (ICMMEd 2020)*, 550(Icmmmed 2020), 257–265. <https://doi.org/10.2991/assehr.k.210508.073>
- Wainer, Howard; I.Braun, H. (2009). *Test Validity (II)*. New York: 365 Broadway Hillsdale, New Jersey.
- Waltz, C. F., & Bausell, B. R. (1981). *Nursing Research: Design Statistics and Computer Analysis*. Davis FA.
- Webb, N. M. (1995). Group Collaboration in Assessment: Multiple Objectives, Processes, and Outcomes. *Educational Evaluation and Policy Analysis*, 17(2), 239–261. <https://doi.org/10.3102/01623737017002239>
- Wegerif, R. (2002). Literature Review in Thinking Skills , Technology and Learning Literature Review in Thinking Skills , Technology and Learning. *FutureLab*, 1–48. Retrieved from [http://www2.futurelab.org.uk/resources/documents/lit\\_reviews/Thinking\\_Skills\\_Review.pdf](http://www2.futurelab.org.uk/resources/documents/lit_reviews/Thinking_Skills_Review.pdf)
- Wei, S., Liu, X., & Jia, Y. (2014). Using Rasch Measurement to Validate the Instrument of Students' Understanding of Models in Science (SUMS). *International Journal of Science and Mathematics Education*, 12(5), 1067–1082. <https://doi.org/10.1007/s10763-013-9459-z>
- Weigle, S. C. (2002). *Assessing Writing*. Ernst Klett Sprachen.
- Weigle, S. C. (2007). Teaching Writing Teachers About Assessment. *Journal of Second Language Writing*, 16(3), 194–209. <https://doi.org/10.1016/j.jslw.2007.07.004>
- White, E. M. (1985). *Teaching and Assessing Writing: Recent Advances in Understanding, Evaluating, and Improving Student Performance*. The Jossey-Bass Higher Education Series. ERIC.
- White, E. M. (2015). Holisticism, 35(4), 400–409.
- Wicking, P. (2019). Formative Assessment of Students From a Confucian Heritage Culture: Insights From Japan. *Assessment and Evaluation in Higher Education*, 45(2), 180–192. <https://doi.org/10.1080/02602938.2019.1616672>
- Wigfi, A. (2009). *Handbook of Motivation at School*. New York: Taylor & Francis or Routledge's.
- Wiliam, D., & Thompson, M. (2007). Integrating Assessment with Learning: What Will It Take to Make It Work? *The Future of Assessment*, 53–82. <https://doi.org/10.4324/9781315086545-3>
- Wilson, L. O. (2016). Anderson and Krathwohl Bloom's Taxonomy Revised Understanding the New Version of Bloom's Taxonomy. *The Second Principle*, 1–8. Retrieved from [https://quincycollege.edu/content/uploads/Anderson-and-Krathwohl\\_Revised-Blooms-Taxonomy.pdf%0Ahttps://thesecondprinciple.com/teaching-essentials/beyond-bloom-cognitive-taxonomy-revised/%0Ahttp://thesecondprinciple.com/teaching-essentials/beyond-bloom-cog](https://quincycollege.edu/content/uploads/Anderson-and-Krathwohl_Revised-Blooms-Taxonomy.pdf%0Ahttps://thesecondprinciple.com/teaching-essentials/beyond-bloom-cognitive-taxonomy-revised/%0Ahttp://thesecondprinciple.com/teaching-essentials/beyond-bloom-cog)
- Wilson, M. (1992). The Ordered artition Model: An Extension of the Partial Credit Model. *Applied Psychological Measurement*, 16(4), 309–325. <https://doi.org/10.1177/014662169201600401>
- Winne, P. H., & Butler, D. L. (1994). Student Cognition in Learning from Teaching. *International Encyclopedia of Education*, 2. Retrieved from [https://scholar.google.co.id/scholar?hl=id&as\\_sdt=0%2C5&q=Winne%2C+P.+H.+%26+Butler%2C+D.+L.+%281994%29.+Student+cognition+in+learning+from+teaching.+In+T.+Husen+%26+T.+Post+lewaite+%28Eds.%29%2C+International+encyclopedia+of+education+%282nd+ed%29.+Oxfo](https://scholar.google.co.id/scholar?hl=id&as_sdt=0%2C5&q=Winne%2C+P.+H.+%26+Butler%2C+D.+L.+%281994%29.+Student+cognition+in+learning+from+teaching.+In+T.+Husen+%26+T.+Post+lewaite+%28Eds.%29%2C+International+encyclopedia+of+education+%282nd+ed%29.+Oxfo)
- WNCP. (2006). *Rethinking Classroom Assessment with Purpose in Mind*.

- Wright, B D, & Stone, M. H. (1979). *Best Test Design*.
- Wright, Benjamin D, & Masters, G. N. (1982). *Rating Scale Analysis*. MESA press.
- Xiao, Y., & Yang, M. (2019). Formative Assessment and Self-Regulated Learning: How Formative Assessment Supports Students' Self-Regulation in English Language Learning. *System*, 81, 39–49. <https://doi.org/10.1016/j.system.2019.01.004>
- Yamada, H. (2000). Comparing “Partial Credit Models” (PCM) and “Rating Scale Models.” *Rasch Measurement Transactions*, (14), 768.
- Yang, F., & Li, F. W. B. (2018). Study on Student Performance Estimation, Student Progress Analysis, and Student Potential Prediction Based on Data Mining. *Computers and Education*, 123(April), 97–108. <https://doi.org/10.1016/j.compedu.2018.04.006>
- Yang, Y., He, P., & Liu, X. (2016). Validation of an Instrument for Measuring Students' Understanding of Interdisciplinary Science in Grades 4-8 over Multiple Semesters: A Rasch Measurement Study. *NARST*, 1–20. Retrieved from <http://wordpress.caset.buffalo.edu/isep/wp-content/uploads/sites/45/2019/12/Yang-and-Liu-NARST-2016.pdf>
- Yeung, S. yin S. (2015). Conception of Teaching Higher Order Thinking: Perspectives of Chinese Teachers in Hong Kong. *Curriculum Journal*, 26(4), 553–578. <https://doi.org/10.1080/09585176.2015.1053818>
- Yildirim, S. G., & Baur, S. W. (2016). Development of Learning Taxonomy for an Undergraduate Course in Architectural Engineering Program. *American Society for Engineering Education*, 1–10. Retrieved from <https://conferences.k-state.edu/aseemidwest2016/files/2016/06/Paper-Yildirim-Baur-11s8avf.pdf>
- You, H. S., Kim, K., Black, K., & Min, K. W. (2018). Assessing Science Motivation for College Students: Validation of The Science Motivation Questionnaire II Using The Rasch-Andrich Rating Scale Model. *Eurasia Journal of Mathematics, Science and Technology Education*, 14(4), 1161–1173. <https://doi.org/10.29333/ejmste/81821>
- Zhang, J. (2012). The Impact of Variability of Item Parameter Estimators on Test Information Function. *Journal of Educational and Behavioral Statistics*, 37(6), 737–757. <https://doi.org/10.3102/1076998612458321>
- Zohar, A., & Alboher Agmon, V. (2018). Raising Test Scores vs. Teaching Higher Order Thinking (HOT): Senior Science Teachers' Views on How Several Concurrent Policies Affect Classroom Practices. *Research in Science and Technological Education*, 36(2), 243–260. <https://doi.org/10.1080/02635143.2017.1395332>
- Zoller, U. (1993). Are Lecture and Learning Compatible? Maybe for LOGS: Unlikely for HOCS. *Journal of Chemical Education*, 70(3), 195–197. <https://doi.org/10.1021/ed070p195>
- Zulyadaini, D. (2017). A Development of Students' Worksheet Based on Contextual Teaching and Learning. *IOSR Journal of Mathematics*, 13(01), 30–38. <https://doi.org/10.9790/5728-1301033038>