



Validitas Isi Instrumen TPACK pada Calon Guru Kimia menggunakan Aplikasi Model Rasch

***Wahyuni Tri Lestari¹, Dwi Maryani Rispatiningsih²**

Sekolah Tinggi Agama Buddha Negeri Raden Wijaya Wonogiri^{1,2}

*Correspondence e-mail; wahyuni.teel@gmail.com

Abstract:

This research aims to assess the content validity of the instrument used to measure Technological Pedagogical Content Knowledge (TPACK) in prospective chemistry teachers. The evaluation encompasses a thorough analysis of the instrument's alignment with the knowledge domains being measured, ensuring that the instrument covers key aspects of TPACK that are relevant to the context of teaching chemistry. Content validity reflects the representation and relevance of a set of items used to measure a concept carried out through rational analysis of the contents of the test through an expert panel assessment. This validity refers to the accuracy of the measurement based on the contents of the instrument to ensure that the scale of the items used has fulfilled the entire concept of the item or the suitability of the item. The instrument used in this study was assessed by nine validators consisting of two chemistry education experts and seven chemistry teachers. The assessment is done by determining the appropriateness or relevance of the test contents to be used through rational analysis by a competent panel or through expert judgment. The instrument validations in this study were 20 multiple choice questions. The assessment is done by giving a score of 1-4 on each item. This research is a quantitative descriptive study using Rasch model analysis. The analysis showed that the exact agreement was 71.4% and the expected agreement was 73.8%. Based on measurements showing relevant items in this research instrument, there are 17, among others number 1, 10, 11, 14, 5, 8, 19, 20, 18, 3, 6, 12, 13, 15, 17, 4 and 9. While there are 3 items that are irrelevant and need to be corrected, including number 16, 2 and 7.

Keywords: Content validity; Rasch model; Item; Validator.

Abstrak:

Penelitian ini bertujuan untuk menilai validitas isi instrumen yang digunakan untuk mengukur TPACK (*Technological Pedagogical Content Knowledge*) pada calon guru kimia. Evaluasi ini mencakup analisis mendalam terhadap kesesuaian instrumen dengan domain pengetahuan yang diukur, memastikan bahwa instrumen mencakup aspek-aspek kunci TPACK yang relevan dengan konteks pengajaran kimia. Validitas isi mencerminkan representasi dan relevansi dari sekumpulan item yang digunakan untuk mengukur sebuah konsep yang dilakukan melalui analisis rasional mengenai isi tes melalui penilaian panel ahli. Validitas ini mengacu pada ketepatan pengukuran didasarkan pada isi (content) instrument untuk memastikan bahwa skala item yang digunakan sudah memenuhi

keseluruhan isi konsep atau kesesuaian item. Instrumen yang digunakan dalam penelitian ini dinilai oleh sembilan validator yang terdiri dari dua ahli pendidikan kimia dan tujuh guru kimia. Penilaian dilakukan dengan menentukan kelayakan atau relevansi isi tes yang digunakan melalui analisis rasional oleh panel yang berkompeten atau melalui expert judgement (penilaian ahli). Validasi instrumen dalam penelitian ini adalah 20 soal pilihan ganda. Penilaian dilakukan dengan memberikan skor 1-4 pada setiap butir soal. Penelitian ini adalah penelitian deskriptif kuantitatif dengan menggunakan analisis model rasch. Hasil analisis menunjukkan bahwa *exact agreement* adalah 71,4% dan *expected agreement* adalah 73,8%. Berdasarkan pengukuran menunjukkan item yang relevan pada instrumen penelitian ini ada 17 yaitu nomer 1, 10, 11, 14, 5, 8, 19, 20, 18, 3, 6, 12, 13, 15, 17, 4 dan 9, sedangkan item yang tidak relevan dan perlu untuk diperbaiki ada 3 yaitu nomer 16, 2 dan 7.

Kata-kata kunci: Item; Model rasch; Validitas isi; Validator

PENDAHULUAN

Validitas merupakan masalah yang sangat penting karena menyangkut ketepatan alat ukur yang akan digunakan. Hal ini dapat dimaknakan bahwa instrument yang tidak tepat atau tidak sesuai maka akan berimplikasi pada validitas hasil riset itu sendiri. Dalam praktiknya, para ahli-ahli telah mengembangkan berbagai cara, teknik maupun metode untuk meningkatkan validitas instrumen ini, dan salah satunya adalah melalui validitas isi sebagai langkah awal untuk penilaian kesesuaian item yang akan digunakan.

Validitas menggambarkan sejauh mana alat ukur (tes) benar-benar mengukur apa yang hendak diukur.¹ Hal yang sama disampaikan oleh Azwar yang mengatakan bahwa validitas berasal dari kata *validity* yang mempunyai arti sejauh mana akurasi suatu tes atau skala dalam menjalankan suatu fungsi pengukurannya.² Secara umum ada tiga pendekatan dalam meneliti validitas suatu alat ukur, yaitu 1) *Content validity* (validitas isi), 2) *Construct validity* (validitas konstruk), 3) *Criterion validity* (validitas kriteria).³ Validitas isi memfokuskan kepada elemen-elemen apa yang ada dalam ukur, sehingga analisis rasional adalah proses utama yang dilakukan dalam analisis validitas isi.⁴ Validitas konstruk memberikan sebuah gambaran yang menunjukkan sejauhmana alat ukur itu menunjukkan

¹ Susan E. Embretson, "Construct Validity: A Universal Validity System or Just Another Test Evaluation Procedure?," *Educational Researcher* 36, no. 8 (2007): 449–455, <https://www.jstor.org/stable/4621099>.

² Azwar S., *Reliabilitas Dan Validitas*, 4th ed. (Yogyakarta: Pustaka Pelajar, 2012).

³ S. Suryabrata, *Pengembangan Alat Ukur Psikologis* (Yogyakarta: Penerbit Andi, 2005).

⁴ Keith Coaley, *An Introduction to Psychological Assessment and Psychometrics* (London: SAGE, 2010), <https://uk.sagepub.com/en-gb/eur/an-introduction-to-psychological-assessment-and-psychometrics/book240294>.

hasil yang sesuai dengan teori.⁵ Validitas kriteria mengaitkan alat ukur dengan alat ukur lain sebagai kriteria, apakah alat ukur itu bisa dijelaskan hasil korelasinya dengan dengan kriterianya berdasarkan teori yang ada.⁶

Validitas isi merupakan validitas yang diestimasikan lewat pengujian terhadap kelayakan atau relevansi isi tes melalui analisis rasional oleh panel yang berkompeten atau melalui *expert judgement* (penilaian ahli) dan biasanya dilakukan oleh tujuh atau lebih ahli.⁷ Validitas isi atau *content validity* memastikan bahwa pengukuran memasukkan sekumpulan item yang memadai dan mewakili yang mengungkap konsep. Atau dengan kata lain, validitas isi merupakan fungsi seberapa baik dimensi dan elemen sebuah konsep yang telah digambarkan.⁸

Secara umum definisi validitas isi adalah sejauh mana elemen-elemen instrument asesmen relevan dan mewakili konstruk alat ukur yang ditargetkan untuk tujuan tertentu.⁹ Validitas isi adalah penjelasan tentang suatu alat ukur secara substantif atau disebut validitas substantif yang fokus kepada konseptualisasi dan sejauhmana konsep-konsep sebelumnya yang ditampilkan dalam kajian *literature*.¹⁰

Mengukur dan melaporkan validitas konten instrumen adalah penting, karena jenis validitas ini juga dapat membantu memastikan validitas konstruk dan memberi kepercayaan kepada pembaca dan peneliti tentang instrument karena melibatkan pakar-pakar untuk memeriksa kelayakan instrument dari sisi konsep dan operasionalisasi.¹¹

⁵ S. Azwar, *Dasar-Dasar Psikometri* (Yogyakarta: Pustaka Pelajar, 2005).

⁶ Robert F. DeVellis and Carolyn T. Thorpe, *Scale Development* (London: Sage Publications, 2003).

⁷ Holli A. Devon et al., "A Psychometric Toolbox for Testing Validity and Reliability," *Journal of nursing scholarship* 39, no. 2 (June 2007): 155–164, accessed November 29, 2023, <https://pubmed.ncbi.nlm.nih.gov/17535316/>.

⁸ U. Sekaran, "Metode Penelitian Bisnis" (Jakarta: Salemba Empat, 2006).

⁹ Stephen N. Haynes, David C.S. Richard, and Edward S. Kubany, "Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods," *Psychological Assessment* 7, no. 3 (1995): 238–247, accessed November 29, 2023, <https://psycnet.apa.org/doiLanding?doi=10.1037%2F1040-3590.7.3.238>.

¹⁰ Lee Anna Clark and David Watson, "Constructing Validity: Basic Issues in Objective Scale Development," *Psychological Assessment* 7, no. 3 (1995): 309–319, <https://psycnet.apa.org/record/1996-93318-001>.

¹¹ Gary Groth-Marnat and A. Jordan Wright, *Handbook of Psychological Assessment*, Terj. Soet. (Yogyakarta: Pustaka Pelajar, 2010).

Item dapat diidentifikasi validitas konten dengan melakukan uji kelayakan instrumen melalui diskusi panel dengan ahli yang kompeten di bidangnya.^{12,13,14} Hasil dari validasi ahli kemudian dianalisis menggunakan model *Rasch*. Model *Rasch* adalah salah satu model pengukuran yang menggunakan model matematika dengan mengubah skor menjadi skor yang dapat diproses untuk memberikan informasi yang lebih valid dan akurat.^{15,16,17} Model *Rasch* juga memungkinkan untuk menyesuaikan variabilitas penilaian yang dilakukan oleh lebih dari satu penilai.^{18,19,20,21} Keuntungan dari model *Rasch* jika dibandingkan dengan teori tes klasik (CTT) adalah bahwa model ini dapat mengidentifikasi jawaban yang relevan atau tidak relevan dari para ahli, identitas penilaian yang tidak tepat, dan memprediksi data yang hilang berdasarkan pola respons sistematis.^{22,23}

METODE PENELITIAN

Metode penelitian ini menggunakan metode deskriptif kuantitatif yang memiliki tujuan untuk menggambarkan peristiwa yang terjadi dengan menggunakan statistik. Data

¹² Laura D. Goodwin and Nancy L. Leech, "The Meaning of Validity in the New Standards for Educational and Psychological Testing:," *Measurement and Evaluation in Counseling and Development* 36, no. 3 (2003): 181–191, accessed November 29, 2023, <https://www.tandfonline.com/doi/abs/10.1080/07481756.2003.11909741>.

¹³ Lewis R. Aiken, "Content Validity and Reliability of Single Items or Questionnaires," *Sage Journals Home* 40, no. 4 (December 1, 1980): 955–959, accessed November 29, 2023, <https://journals.sagepub.com/doi/10.1177/001316448004000419>.

¹⁴ Suzanne Lane, "Validity of High-Stakes Assessment: Are Students Engaged in Complex Thinking?," *Educational Measurement: Issues and Practice* 23, no. 3 (September 1, 2004): 6–14, <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.2004.tb00160.x>.

¹⁵ Purya Baghaei and Nazila Amrahi, "Validation of a Multiple Choice English Vocabulary Test with the Rasch Model," *Journal of Language Teaching and Research* 2, no. 5 (2011): 1052–1060.

¹⁶ B. Umintono and W. Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan* (Cimahi: Trim Komunikata Publishing House, 2015).

¹⁷ B. Sumintono and W. Widhiarso, *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial* (Cimahi: Trim Komunikata Publishing House, 2014).

¹⁸ Thanh Dong Nguyen and David Ng Foo Seong, "Applying the Rasch Model to Investigate Singapore Principals' Instructional Leadership Practices," *Leading and Managing* 20, no. 2 (2014): 1–26, accessed November 29, 2023, <https://search.informit.org/doi/abs/10.3316/aeipt.206326>.

¹⁹ Sarah Goodwin, "A Many-Facet Rasch Analysis Comparing Essay Rater Behavior on an Academic English Reading/Writing Test Used for Two Purposes," *Assessing Writing* 30 (October 1, 2016): 21–31.

²⁰ Komarudin Kudiya et al., "Batik Artisans' Judgment of Batik Wax Quality and Its Criteria: An Application of the Many-Facets Rasch Model," *Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings* (2018): 27–37, https://link.springer.com/chapter/10.1007/978-981-10-8138-5_3.

²¹ Richard R. Sudweeks, Suzanne Reeve, and William S. Bradshaw, "A Comparison of Generalizability Theory and Many-Facet Rasch Measurement in an Analysis of College Sophomore Writing," *Assessing Writing* 9, no. 3 (January 1, 2004): 239–261.

²² Andrea Benešová and Jiří Tupa, "Requirements for Education and Qualification of People in Industry 4.0," *Procedia Manufacturing* 11 (January 1, 2017): 2195–2202.

²³ Tritiyatma Hadinugrahaningsih, Yuli Rahmawati, and Achmad Ridwan, "Developing 21st Century Skills in Chemistry Classrooms: Opportunities and Challenges of STEAM Integration," *AIP Conference Proceedings* 1868, no. 1 (August 4, 2017): 30008, [/aip/acp/article/1868/1/030008/641094/Developing-21st-century-skills-in-chemistry](https://doi.org/10.1063/1.5000008).

yang digunakan dalam penelitian ini diperoleh dari hasil penilaian instrumen yang telah dilakukan oleh sembilan validator yang terdiri dari dua ahli pendidikan kimia dan tujuh guru kimia. Validasi instrumen dalam penelitian ini adalah 20 soal pilihan ganda. Penilaian dilakukan dengan memberikan skor 1-4 pada setiap butir soal. Nilai 1-2-3-4 terletak di seluruh item, masing-masing angka diberi arti sebagai; tidak relevan, kurang relevan, cukup relevan, relevan. Analisis data menggunakan teknik analisis deskriptif kuantitatif. Penelitian ini menggunakan *software minifac 3.82.3* yang dikembangkan oleh Linacre.²⁴ Perangkat lunak yang digunakan dalam penelitian ini adalah *minifac*. *Minifac* adalah program komputer khusus yang bekerja di bawah sistem *Microsoft Windows* yang dibuat oleh John Linacre. Perangkat ini adalah program pengujian yang dapat melibatkan banyak penilai. *Minifac* yang dikembangkan oleh Linacre dapat melibatkan interaksi antara validator, kualitas, dan item.

HASIL DAN PEMBAHASAN

Validitas konten merupakan validitas yang diestimasi lewat pengujian terhadap kelayakan atau relevansi isi tes melalui analisis rasional oleh panel yang berkompeten atau melalui *expert judgement* (penilaian ahli).²⁵ Tujuan validasi isi adalah mengurangi variasi potensi kesalahan dalam pembuatan instrumen dan meningkatkan kemungkinan diperolehnya indeks validitas konstruk dalam studi lanjutan.²⁶ Dalam validasi isi, peneliti merevisi kesalahan kata-kata yang ada pada item sesuai dengan masukan dari para ahli. Sembilan ahli setuju bahwa item yang dikembangkan memiliki validitas isi yang dapat diterima.

Data kuantitatif yang diperoleh dalam penelitian kemudian dianalisis dengan model *Rasch* menggunakan *software minifac* yang dikembangkan oleh Linacre.²⁷ Pengukuran model *Rasch* mentransformasikan hasil pengukuran yang diperoleh untuk setiap sumber variabilitas menjadi skala interval nyata yang diketahui sebagai "Skala Logit". Selain estimasi log untuk setiap sisi, Model *Rasch* juga memungkinkan penentuan kualitas untuk

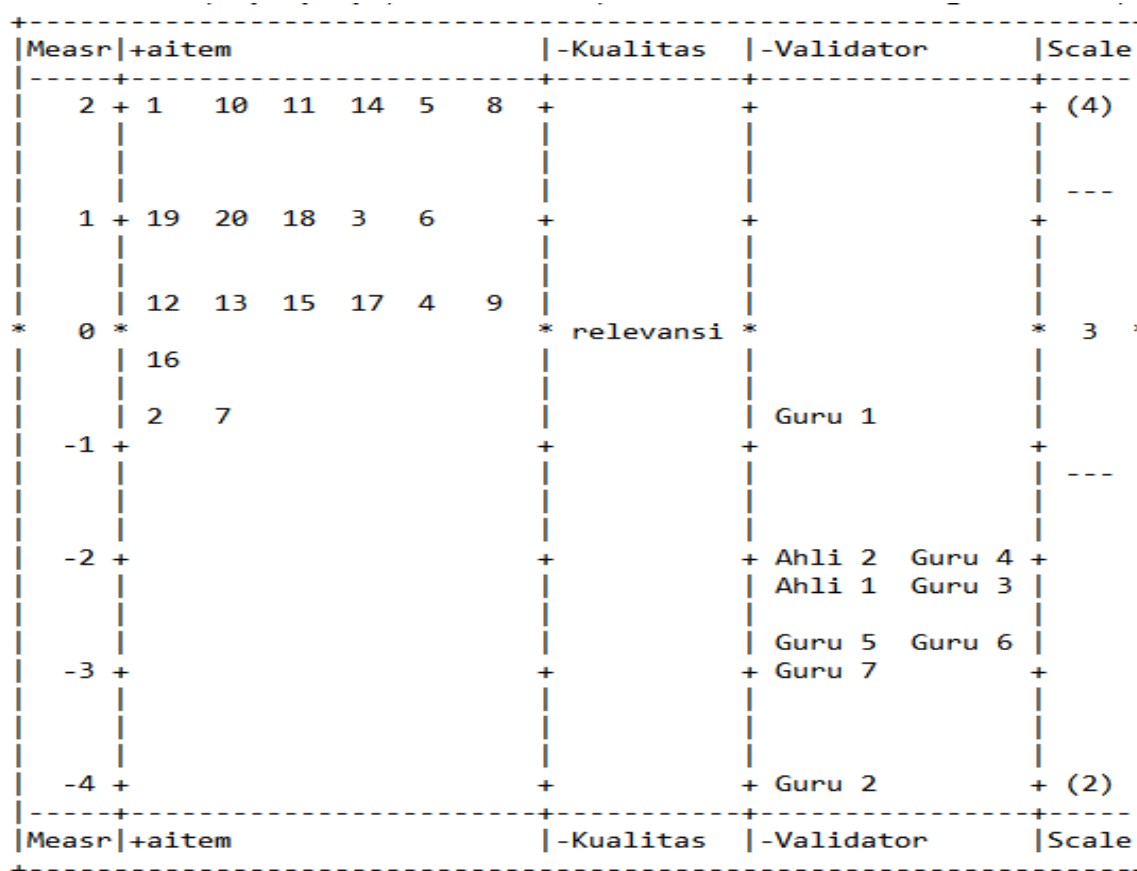
²⁴ J. M. Linacre, *FACETS Version 3.82.3 [Computer Software and Manual]* (Chicago: Winsteps.com, 2013).

²⁵ Sekaran, "Metode Penelitian Bisnis."

²⁶ Haynes, Richard, and Kubany, "Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods."

²⁷ Linacre, *FACETS Version 3.82.3 [Computer Software and Manual]*.

setiap item pada instrument.²⁸ Hasil pengukuran grafis validasi yang dilakukan oleh ahli menggunakan *software minifac* seperti pada Gambar 1. berikut:



Gambar 1. Hasil pengukuran grafis menggunakan *software minifac*

Gambar 1. Menunjukkan hasil dari pengukuran secara grafis dengan menggunakan *software minifac*. Gambar ini menunjukkan antara lain: skala logit, item, kualitas dan validator. Kolom disebelah kiri menunjukkan skala logit, skala logit adalah skala dengan interval yang sama dan bersifat linear yang berasal dari data ratio (*odds ratio*) dan bukannya data mentah skor yang didapat.^{29,30}

Kolom kedua menunjukkan adanya distribusi item dari hasil pengukuran dan kolom ketiga berupa kualitas (relevan atau tidak) item yang telah dibuat. Distribusi item menunjukkan adanya kualitas dari masing-masing item yang diperoleh dari hasil pengukuran. Item yang berada sejajar atau diatas kualitas relevansi menunjukkan bahwa item tersebut relevan, item yang relevan dari hasil pengukuran menurut validator pada

²⁸ Andrea Révész, “Working Memory and the Observed Effectiveness of Recasts on Different L2 Outcome Measures,” *Language Learning* 62, no. 1 (March 1, 2012): 93–132, accessed November 29, 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9922.2011.00690.x>.

²⁹ Umintono and Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*.

³⁰ Sumintono and Widhiarso, *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*.

instrumen penelitian ini ada 17 item yaitu nomer 1, 10, 11, 14, 5, 8, 19, 20, 18, 3, 6, 12, 13, 15, 17, 4 dan 9, sedangkan item yang berada dibawah kualitas relevan menunjukkan bahwa item tersebut tidak relevan dan perlu untuk diperbaiki, dari hasil pengukuran menurut validator item yang tidak relevan ada 3 item yaitu nomor 16, 2 dan 7. Item yang tidak relevan perlu adanya perbaikan sesuai dengan saran yang telah diberikan oleh validator.

Kolom keempat menunjukkan validator atau yang memberi penilaian pada item yang digunakan dalam penelitian ini. Validator pada penelitian ini berjumlah 9 orang diantaranya 2 ahli pendidikan kimia dari Universitas Sebelas Maret Surakarta dan 7 guru kimia dari SMA Negeri/Swasta favorit di Solo. Validator yang memberi penilaian yang relatif “pelit” berada paling atas (logit tertinggi), dalam penelitian ini adalah guru 1. Validator yang memberikan penilaian paling rendah (logit terendah) adalah validator yang memberikan penilaian paling murah diantara validator yang lain, dalam penelitian ini adalah guru 2. Validator yang lain mengelompokkan antara logit -2 ke -3, yang berarti bahwa mereka memiliki pandangan yang hampir sama tentang item yang sedang dinilai.

Hasil analisis pengukuran menggunakan *software minifac* dari validator dapat dilihat pada tabel 1.

Tabel 1. Ringkasan hasil analisis pengukuran menggunakan *software minifac*

<i>exact agreement</i>	<i>expected agreement</i>
71,4%	73.8%

Analisis dengan menggunakan *software minifac* menunjukkan bahwa penilaian ahli (*exact agreement*) yang diperoleh adalah 71,4% sedangkan penilaian yang diharapkan dengan menggunakan model *Rasch* (*expected agreement*) 73,8%, Hal ini menunjukkan bahwa hasil penilaian dari validator tidak jauh berbeda dari hasil penilaian yang diharapkan dengan menggunakan model *Rasch*.

KESIMPULAN

Validitas konten ini dilakukan dengan menilai instrumen melalui diskusi panel oleh para ahli yang kompeten di bidangnya. Instrumen yang dinilai dalam penelitian ini berjumlah 20 soal pilihan ganda yang divalidasi oleh sembilan validator, yaitu 2 ahli pendidikan kimia dan 7 guru kimia. validasi dilakukan dengan memberikan skor 1-4 pada setiap item. Hasil dari data validasi kemudian dianalisis menggunakan model *Rasch* melalui *software minifac*. Analisis menunjukkan bahwa penilaian ahli (*exact agreement*) yang diperoleh adalah 71,4% sedangkan penilaian yang diharapkan dengan menggunakan model *Rasch* (*expected*

agreement) 73,8%. Hal ini menunjukkan bahwa hasil penilaian validator tidak jauh berbeda dari hasil penilaian yang diharapkan menggunakan perhitungan *Rasch*. Jika dilihat dari sisi validator, kategori validator yang paling ketat adalah guru 1, sedangkan validator paling mudah adalah guru 2. Validator yang lain mengelompokkan antara logit -2 ke -3, yang berarti bahwa mereka memiliki pandangan yang hampir sama tentang item yang sedang dinilai. Berdasarkan pengukuran menunjukkan bahwa item yang relevan pada instrumen penelitian ini ada 17 item yaitu nomor 1, 10, 11, 14, 5, 8, 19, 20, 18, 3, 6, 12, 13, 15, 17, 4 dan 9, sedangkan item yang tidak relevan dan perlu untuk diperbaiki ada 3 item yaitu nomor 16, 2 dan 7.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM), Universitas Sebelas Maret, Surakarta, Indonesia untuk menyediakan dana hibah untuk penelitian melalui Kelompok Penelitian 2019.

REFERENSI

- Aiken, Lewis R. "Content Validity and Reliability of Single Items or Questionnaires." *Sage Journals Home* 40, no. 4 (December 1, 1980): 955–959. Accessed November 29, 2023. <https://journals.sagepub.com/doi/10.1177/001316448004000419>.
- Azwar, S. *Dasar-Dasar Psikometri*. Yogyakarta: Pustaka Pelajar, 2005.
- Baghaei, Purya, and Nazila Amrahi. "Validation of a Multiple Choice English Vocabulary Test with the Rasch Model." *Journal of Language Teaching and Research* 2, no. 5 (2011): 1052–1060.
- Benešová, Andrea, and Jiří Tupa. "Requirements for Education and Qualification of People in Industry 4.0." *Procedia Manufacturing* 11 (January 1, 2017): 2195–2202.
- Clark, Lee Anna, and David Watson. "Constructing Validity: Basic Issues in Objective Scale Development." *Psychological Assessment* 7, no. 3 (1995): 309–319. <https://psycnet.apa.org/record/1996-93318-001>.
- Coaley, Keith. *An Introduction to Psychological Assessment and Psychometrics*. London: SAGE, 2010. <https://uk.sagepub.com/en-gb/eur/an-introduction-to-psychological-assessment-and-psychometrics/book240294>.
- DeVellis, Robert F., and Carolyn T. Thorpe. *Scale Development*. London: Sage Publications, 2003.
- Devon, Holli A., Michelle E. Block, Patricia Moyle-Wright, Diane M. Ernst, Susan J.

- Hayden, Deborah J. Lazzara, Suzanne M. Savoy, and Elizabeth Kostas-Polston. "A Psychometric Toolbox for Testing Validity and Reliability." *Journal of nursing scholarship* 39, no. 2 (June 2007): 155–164. Accessed November 29, 2023. <https://pubmed.ncbi.nlm.nih.gov/17535316/>.
- Goodwin, Laura D., and Nancy L. Leech. "The Meaning of Validity in the New Standards for Educational and Psychological Testing:" *Measurement and Evaluation in Counseling and Development* 36, no. 3 (2003): 181–191. Accessed November 29, 2023. <https://www.tandfonline.com/doi/abs/10.1080/07481756.2003.11909741>.
- Goodwin, Sarah. "A Many-Facet Rasch Analysis Comparing Essay Rater Behavior on an Academic English Reading/Writing Test Used for Two Purposes." *Assessing Writing* 30 (October 1, 2016): 21–31.
- Groth-Marnat, Gary, and A. Jordan Wright. *Handbook of Psychological Assessment*. Terj. Soet. Yogyakarta: Pustaka Pelajar, 2010.
- Hadinugrahaningsih, Tritiyatma, Yuli Rahmawati, and Achmad Ridwan. "Developing 21st Century Skills in Chemistry Classrooms: Opportunities and Challenges of STEAM Integration." *AIP Conference Proceedings* 1868, no. 1 (August 4, 2017): 30008. [/aip/acp/article/1868/1/030008/641094/Developing-21st-century-skills-in-chemistry](https://aip.acp/article/1868/1/030008/641094/Developing-21st-century-skills-in-chemistry).
- Haynes, Stephen N., David C.S. Richard, and Edward S. Kubany. "Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods." *Psychological Assessment* 7, no. 3 (1995): 238–247. Accessed November 29, 2023. <https://psycnet.apa.org/doiLanding?doi=10.1037%2F1040-3590.7.3.238>.
- Kudiya, Komarudin, Bambang Sumintono, Setiawan Sabana, and Agus Sachari. "Batik Artisans' Judgment of Batik Wax Quality and Its Criteria: An Application of the Many-Facets Rasch Model." *Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings* (2018): 27–37. https://link.springer.com/chapter/10.1007/978-981-10-8138-5_3.
- Lane, Suzanne. "Validity of High-Stakes Assessment: Are Students Engaged in Complex Thinking?" *Educational Measurement: Issues and Practice* 23, no. 3 (September 1, 2004): 6–14. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.2004.tb00160.x>.
- Linacre, J. M. *FACETS Version 3.82.3 [Computer Software and Manual]*. Chicago: Winsteps.com, 2013.
- Nguyen, Thanh Dong, and David Ng Foo Seong. "Applying the Rasch Model to Investigate

- Singapore Principals' Instructional Leadership Practices.” *Leading and Managing* 20, no. 2 (2014): 1–26. Accessed November 29, 2023. <https://search.informit.org/doi/abs/10.3316/aeipt.206326>.
- Révész, Andrea. “Working Memory and the Observed Effectiveness of Recasts on Different L2 Outcome Measures.” *Language Learning* 62, no. 1 (March 1, 2012): 93–132. Accessed November 29, 2023. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9922.2011.00690.x>.
- S., Azwar. *Reliabilitas Dan Validitas*. 4th ed. Yogyakarta: Pustaka Pelajar, 2012.
- Sekaran, U. “Metode Penelitian Bisnis.” Jakarta: Salemba Empat, 2006.
- Sudweeks, Richard R., Suzanne Reeve, and William S. Bradshaw. “A Comparison of Generalizability Theory and Many-Facet Rasch Measurement in an Analysis of College Sophomore Writing.” *Assessing Writing* 9, no. 3 (January 1, 2004): 239–261.
- Sumintono, B., and W. Widhiarso. *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. Cimahi: Trim Komunikata Publishing House, 2014.
- Suryabrata, S. *Pengembangan Alat Ukur Psikologis*. Yogyakarta: Penerbit Andi, 2005.
- Susan E. Embretson. “Construct Validity: A Universal Validity System or Just Another Test Evaluation Procedure?” *Educational Researcher* 36, no. 8 (2007): 449–455. <https://www.jstor.org/stable/4621099>.
- Umintono, B., and W. Widhiarso. *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Cimahi: Trim Komunikata Publishing House, 2015.