



Penggunaan Rasch Model untuk Validasi *Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI)*

Application of the Rasch Model for Validating Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI)

Sukmawati Ira Saputri^{1*}, Yusfa Lestari², Nurqadriyanti Hasanuddin³, Salizmi Zulfi Maulidita⁴

¹PGSD, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Riau

²Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sulawesi Barat

³Pendidikan Bahasa Inggris, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sulawesi Barat

⁴Teaching Korean as Second Language, Faculty of Humanities, Chosun University

*Correspondence e-mail: yusfa.lestari@unsulbar.ac.id

Received:

Revised:

Accepted :

Published:

Abstrak

Penelitian ini bertujuan untuk menganalisis validitas dan reliabilitas dari hasil adopsi *Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI)*. Penelitian ini merupakan penelitian kuantitatif dengan desain survey yang memiliki subjek penelitian adalah 69 orang calon guru kimia di sebuah Universitas di Indonesia. Sebelum digunakan, instrumen diterjemahkan dari Bahasa Inggris ke Bahasa Indonesia dengan metode *back translation*. Data hasil penelitian selanjutnya dianalisis menggunakan Winstep dan didapatkan hasil validitas dan reliabilitas yang baik. RSCRDI memiliki *Cronbach alpha* 0,65, *person reliability* 0,60 dan *item reliability* 0,76. Hasil validasi kontruk juga menunjukkan hasil yang baik ditinjau dari *mean square* (MNSQ), *Z-standard* (ZSTD) dan *point measure correlation* (Pt Mea Corr). Hasil penelitian ini menunjukkan bahwa RSCRDI dapat digunakan sebagai instrumen pengukur konsep bagi calon guru kimia di Indonesia yang valid dan reliabel.

Kata Kunci

Rasch model, instrumen, reaksi kimia, calon guru kimia

Abstract

This study seeks to evaluate the validity and reliability of the outcomes derived from the adoption of the *Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI)*. This research employs a quantitative, survey-based approach, involving 69 prospective chemistry teachers at a university in Indonesia. Prior to implementation, the instrument was translated from English to Indonesian using the back-translation method. The collected data were subsequently

analyzed using Winstep, yielding strong validity and reliability results. The instrument exhibited a Cronbach's alpha of 0.65, a person reliability of 0.60, and an item reliability of 0.76. Furthermore, the construct validation results were favorable, as evidenced by mean score (MNSQ), Z-standard (ZSTD), and point measure correlation (Pt Mea Corr) indices. These findings indicate that the RSCEDI is a valid and reliable instrument for assessing conceptual understanding among prospective chemistry teachers in Indonesia.

Keywords

Rasch model, instrument, chemical reactions, pre-service teacher

PENDAHULUAN

Kimia merupakan salah satu bagian dari ilmu sains yang penting dipelajari oleh siswa agar mereka dapat menjadi generasi emas yang diharapkan Indonesia (Ramadhana & Qudratuddarsi, 2024). Kimia dianggap sebagai mata pelajaran yang sulit dipelajari oleh siswa atau diajarkan oleh guru. Mata pelajaran ini pada hakikatnya sangat abstrak, konseptual dan diikuti oleh perhitungan matematis yang membutuhkan keterampilan tinggi dalam proses pembelajaran (Wei, Liu, Wang, & Wang, 2012). Salah satu materi yang sulit dipelajari adalah reaksi kimia dan materi tersebut menjadi prasyarat untuk mempelajari topik-topik lain dalam ilmu kimia. Memahami reaksi kimia mengharuskan siswa untuk menjelaskan konsep dan fenomena pada tiga tingkat representasi (Chandrasegaran, Treagust, & Mocerino, 2011; Santos & Arroio, 2016). Tingkat representasi pertama dimulai dari apa yang siswa lihat di sekitar mereka (representasi makroskopis: yang dapat diamati, disentuh, dan dicium) (Taber, 2013), tingkat berikutnya adalah apa yang dapat mereka jelaskan tentang peristiwa tersebut dari perspektif kimia (tingkat sub-mikroskopis: atom, molekul, ion, dan struktur) (Rappoport & Ashkenazi, 2008), dan tingkat terakhir adalah bagaimana mereka mengomunikasikan menggunakan rumus kimia atau persamaan kimia (representasi simbolik: simbol, rumus, persamaan, molaritas, manipulasi matematis, dan grafik) (Taber, 2013; Touli, Talbi, & Radid, 2015).

Ketika siswa diminta untuk menggambarkan fenomena kimia, mereka sering kali kesulitan untuk menghubungkan ketiga tingkat representasi tersebut dan terjadi miskonsepsi (Gurel, Eryilmaz, & McDermott, 2015). Miskonsepsi adalah pandangan siswa tentang fenomena kimia yang bertentangan dengan pandangan para ahli kimia (Kirbulut & Geban, 2014; Romine, Todd, & Clark, 2016). Sering kali, para guru memandang miskonsepsi baik sebagai hambatan. Mereka percaya bahwa miskonsepsi menghalangi pembelajaran dengan mencegah akses ke ide-ide ilmiah utama, menghambat kemampuan siswa untuk memahami konsep, dan mempengaruhi cara siswa memperoleh pengetahuan baru. Oleh karena itu, guru mencari cara untuk mendeteksi miskonsepsi ini agar dapat diatasi, diubah, diganti, dihindari, dan dihilangkan (Schultz et al., 2017). Deteksi miskonsepsi sangat menguntungkan karena hasilnya dapat menjadi rekomendasi untuk memodifikasi metode pengajaran dan mengukur peningkatan pembelajaran (menilai kemajuan atau efektivitas pengajaran) jika ada pelaksanaan pre-test dan post-test (Liu, 2010).

Identifikasi miskonsepsi dapat dilakukan dengan menggunakan instrumen *two-tier multiple choice questions (TTMCQ)*. *Two-Tier Multiple Choice Questions (TTMCQ)* adalah jenis soal pilihan ganda yang terdiri dari dua tingkatan. Tingkat pertama berisi soal pilihan ganda standar, sementara tingkat kedua meminta siswa untuk memberikan alasan atau justifikasi atas jawaban mereka di tingkat pertama. TTMCQ digunakan untuk mengukur pemahaman konsep secara lebih mendalam, bukan hanya sekadar ingatan fakta (Laliyo, Hamdi, Pikoli, Abdullah, & Panigoro, 2021). Instrumen ini sudah banyak dikembangkan namun perlu ditinjau lagi validitas dan reliabilitasnya setelah dilakukan penerjemahan ke bahasa Indonesia. Adopsi instrumen juga merupakan langkah yang strategis untuk mendapatkan instrumen yang valid dan reliabel dengan tahapan yang lebih mudah dan efisien. Dalam validasi ini, peneliti menggunakan Rasch model dari *Item Response*

Theory yang berbeda dengan peneliti sebelumnya yang menggunakan *Classical Test Theory* dalam menganalisis validitas dan reliabilitas instrumen yang digunakan. Instrumen yang digunakan dalam penelitian ini adalah *Representational Systems and Chemical Reactions Diagnostic Instrument (RSCRDI)*.

Adapun rumusan masalah dalam penelitian ini adalah

- 1) Berapa nilai reliabilitas dan *separation* dari RSCRDI?
- 2) Apakah setiap butir pertanyaan pada RSCRDI memenuhi *goodness of model fit* dari Rasch model?

METODE

Penelitian ini merupakan penelitian kuantitatif dengan desain survey yang bertujuan untuk mendapatkan data-data dari responden dalam bentuk angka yang akan diolah untuk mencapai tujuan penelitian (Hidayat et al., 2024; Qudratuddarsi et al., 2019). Tahapan penelitian ini dimulai dengan memilih instrumen yang akan diadopsi kemudian meminta izin kepada pengembang instrumen untuk menggunakan instrumen tersebut (Hidayat et al., 2021). Pemilihan instrumen ini didasarkan pada beberapa pertimbangan: (1) bentuk instrumen berupa soal pilihan ganda dua tingkat yang dianggap efektif untuk mendeteksi miskonsepsi karena dapat menggali pemahaman siswa lebih mendalam dibandingkan dengan soal pilihan ganda biasa. (2) Penggunaan yang luas dikarenakan memiliki keunggulan dibanding tes subjektif (seperti wawancara, esai, dan tes terbuka) maupun tes objektif (seperti soal pilihan ganda) (Lin, 2016).

Instrumen tersebut selanjutnya diterjemahkan ke dalam Bahasa Indonesia melalui metode *back translation* (Hidayat et al., 2022). Dalam proses penerjemahan, peneliti melibatkan pakar bahasa dan pakar kimia yang fasih berbahasa Inggris. Sebelum pengumpulan data, responden diinformasikan bahwa tes tersebut adalah tes diagnostik dan hasilnya tidak akan mempengaruhi nilai mereka. Pada penelitian ini, peneliti menggunakan tes berbasis kertas dan pensil yang memungkinkan peneliti mengamati proses pengumpulan data, memperoleh tingkat respons yang lebih baik, serta memastikan keterjangkauan responden (Zuidgeest, Hendriks, Koopman, Spreeuwenberg, & Rademakers, 2011; Hidayat, Idris, et al., 2021)). Adapun statistik deskriptif responden terdapat pada Tabel 1.

Karaktestik	Jumlah (n)	Persentase (%)
Jenis kelamin		
Laki-laki	10	14,49%
Perempuan	59	85,81%
Tahun perkuliahan		
Tahun pertama	27	39.13%
Tahun kedua	21	30.43%
Tahun ketiga	21	30.43%
Jumlah total	69	100%

Tabel 1. Statistik deskriptif responden

Data hasil penelitian selanjutnya dilakukan skoring berdasarkan rubrik pada Tabel 2 dan ditabulasikan pada Ms Excel 2019 dan dianalisis menggunakan Winstep 3.7.3. Data *goodness of model fit* juga dievaluasi berdasarkan Boone, Staver, dan Yale (2014) dengan kriteria evaluasi kesesuaian model adalah sebagai berikut: (a) rentang yang diterima untuk nilai *infit* dan *outfit mean square* (MNSQ) adalah $0,5 < \text{MNSQ} < 1,5$; (b) rentang yang ditoleransi untuk nilai *infit* dan *outfit Z-Standard* (ZSTD) adalah $-2,0 < \text{ZSTD} < +2,0$; dan (c) nilai yang diterima untuk *Point measure correlation* (Pt Mean Corr) harus bernilai positif.

Pertanyaan	Alasan	Skor
Pola jawaban	Benar	Benar

Pola jawaban	Benar	Salah
Pola jawaban	Salah	Benar
Pola jawaban	Salah	Salah
Pola jawaban	Benar	Benar

Tabel 2. Rubrik Skoring (Referensi: Park & Liu, 2019)

HASIL

Validitas dan reliabilitas instrumen RSCRDI ditinjau dari *Item Response Theory* pada penelitian ini difokuskan pada *Cronbach's alpha* (0,65), *person reliability* (0,60), *item reliability* (0,76), *person separation* (1,23) dan *item separation* (1,80). Dalam penelitian ini, *person reliability* memiliki nilai dibawah 0,65. Nilai berikutnya yang perlu dipertimbangkan adalah separation. Berdasarkan Sumintono dan Widhiarso (2015), salah satu persamaan untuk memperkirakan separation item adalah $H(separation) = \{(4 \times separation) + 1\} / 3 = 2,733$, atau 3. Ini berarti bahwa item dapat membedakan kemampuan responden menjadi tinggi, sedang, dan rendah. Demikian pula, separation yang lebih tinggi dari satu menunjukkan penyebaran item dan person yang baik (Chan, Ismail, & Sumintono, 2015; Gracia, 2005).

	Nilai
<i>Cronbach's alpha</i>	0,65
<i>Person Reliability</i>	0,60
<i>Item Reliability</i>	0,76
<i>Person separation</i>	1,23
<i>Item separation</i>	1,80

Tabel 3. Reliabilitas dan Separation pada RSCRDI

Secara umum, setiap butir pertanyaan pada RSCRDI memiliki hasil *model fit* dalam rentang yang dapat diterima sesuai data pada Tabel 4. Walaupun ada beberapa item yang memiliki index yang berada diluar standar yang ditentukan, butir soal tersebut tetap dipertahankan karena indikator lainnya terpenuhi.

Item	Infit		Outfit		Pt Mea Corr
	MNSQ	ZSTD	MNSQ	ZSTD	
1	0,8	-1,3	0,88	-0,6	0,38
2	1,19	1,5	1,21	-0,6	0,35
3	0,95	-0,3	0,89	-0,6	0,58
4	0,76	-2	0,77	-1,8	0,38
5	0,93	-0,5	0,91	-0,6	0,55
6	0,75	-2,1*	0,76	-1,8	0,43
7	1,17	1,3	1,16	1,2	0,48
8	1,08	0,7	1,07	0,5	0,32
9	1,21	1,7	1,18	1,4	0,34
10	0,98	-0,1	0,96	-0,2	0,41
11	0,94	-0,4	0,93	-0,4	0,56
12	1,05	0,5	1,06	0,5	0,41
13	1,15	1,2	1,22	1,6	0,14
14	0,95	-0,3	0,91	-0,6	0,49
15	1,02	0,2	1,05	0,4	0,32
Rata-rata	0,99	0,01	0,99	0,03	0,41
Standar Deviasi	0,15	1,19	0,15	1,10	0,11
Nilai Maksimal	0,75	-2,1	0,76	-1,8	0,14
Nilai Minimal	1,21	1,7	1,22	1,6	0,58

Tabel 4. Hasil dari Model Fit

PEMBAHASAN

Rasch model adalah salah satu model dalam *Item Response Theory (IRT)* yang digunakan untuk menganalisis data dalam pengukuran pendidikan dan psikologi. Prinsip dasar model ini adalah probabilitas seorang individu menjawab benar suatu item tergantung pada dua faktor: tingkat kemampuan individu dan tingkat kesulitan item tersebut. *Rasch model* mengasumsikan bahwa jika seorang individu memiliki kemampuan yang lebih tinggi daripada tingkat kesulitan item, maka probabilitas mereka menjawab benar akan lebih tinggi. Sebaliknya, jika kemampuan individu lebih rendah dari kesulitan item, maka probabilitas menjawab benar akan lebih rendah. Model ini juga mengasumsikan bahwa hubungan antara kemampuan dan kesulitan ini bersifat logistik, yang berarti perubahan dalam kemampuan individu akan berdampak pada perubahan probabilitas menjawab benar secara konsisten (Qudratuddarsi et. Al., 2022).

Hasil penelitian validasi RSCREDI menunjukkan bahwa reliabilitas person adalah 0,6 dan berada di bawah skor yang diharapkan. Dalam bidang instrumen diagnostik, beberapa penelitian (misalnya, Caleon & Subramaniam, 2010; Hoe & Subramaniam, 2016; Sreenivasulu & Subramaniam, 2013, 2014; Yan & Subramaniam, 2018) yang dipublikasikan dalam artikel-artikel yang bereputasi tinggi juga menemukan hasil reliabilitas yang tidak memuaskan, dengan nilai reliabilitas yang lebih rendah dari 0,5 (nilai minimum adalah 0,15). Hasil tersebut dipublikasikan di *Chemistry Education Research and Practice*, *International Journal of Science Education*, dan *Research in Science Education*. Ada beberapa hal yang mungkin menjadi penyebab rendahnya reliabilitas antara lain 1) Pemahaman kognitif sering kali bersifat kompleks dan bervariasi antar individu. Respon peserta terhadap item kognitif bisa sangat dipengaruhi oleh faktor-faktor seperti latar belakang pengetahuan, strategi pemecahan masalah yang digunakan, dan tingkat motivasi. Variabilitas ini dapat mengurangi konsistensi skor dari waktu ke waktu. 2) Item kognitif biasanya dirancang untuk mengukur konsep abstrak dan kemampuan berpikir yang lebih tinggi, seperti analisis, sintesis, dan evaluasi. Ketika item-item ini tidak dirancang dengan cermat, mereka mungkin tidak menangkap secara konsisten aspek yang ingin diukur, sehingga mengurangi reliabilitas instrumen.

Hasil analisis menunjukkan bahwa setiap item sesuai dengan model pengukuran Rasch. Hanya butir soal no. 6 yang tidak memenuhi kriteria *infitt Z-Standard (ZSTD)* yang ditoleransi. Nilai *ZSTD* yang tinggi pada Item 6 menunjukkan bahwa item tersebut gagal membedakan antara siswa dalam hal kemampuan yang ditargetkan, sehingga kemungkinan mengukur konstruksi yang berbeda dari item lainnya. Nilai *outfit* yang besar jelas lebih bermasalah dibandingkan dengan nilai *outfit* yang kecil (Liu, Lee, Linn, & Liu, 2011). Meskipun terdapat beberapa nilai di luar rentang yang dapat diterima, terutama pada *infitt* dan *outfit Z-Standard (ZSTD)*, nilai-nilai tersebut tidak muncul secara bersamaan. Oleh karena itu, semua item dianggap sebagai item yang baik untuk mengukur pemahaman konseptual siswa (Sumintono & Widhiarso, 2015).

Penggunaan Model Rasch dalam instrumen diagnostik aspek kognitif telah banyak diterapkan dalam bidang pendidikan dan psikologi untuk memastikan keakuratan dan keandalan pengukuran kemampuan kognitif individu. Dalam konteks diagnostik, Model Rasch memungkinkan pendeteksian butir-butir soal yang tidak berfungsi dengan baik atau yang tidak sesuai dengan model, yang dapat mengindikasikan adanya bias atau kesalahan dalam pengukuran. Hal ini sangat penting dalam pengembangan instrumen diagnostik yang bertujuan untuk mengidentifikasi kemampuan atau keterbatasan kognitif individu secara akurat.

KESIMPULAN

Penelitian ini menunjukkan bahwa RSCREDI versi terjemahan memiliki validitas dan reliabilitas yang baik untuk digunakan sebagai instrumen untuk mengukur kemampuan konsep calon guru kimia pada materi representasi

dan reaksi kimia. Instrumen ini memiliki Cronbach alpha 0,65, person reliability 0,60 dan item reliability 0,76. Hasil validasi kontruk juga menunjukkan hasil yang bagus ditinjau dari *mean square (MNSQ)*, *Z score (ZSTD)* dan *point measure correlation (Pt Mea Corr)*. Penelitian ini dapat menjadi rujukan bagi peneliti bagaimana mengadopsi instrumen *two-tier multiple choice* agar dapat digunakan untuk menganalisis berbagai pengetahuan siswa maupun calon guru di Indonesia.

Referensi

- Amin, M. 2017. Biologi sebagai Sumber Belajar untuk Generasi Mendatang yang Berintegritas Dan Berperadaban Tinggi. Pidato Pengukuhan Jabatan Guru Besar. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Negeri Malang. Kemristekdikti
- Boone, W.J, Staver, J. ., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer Netherlands
- Caleon, I. S., & Subramaniam, R. (2010). Do students know What they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, 40(3), 313–337. <https://doi.org/10.1007/s11165-009-9122-4>
- Chan, S. W., Ismail, Z., & Sumintono, B. (2015). The impact of statistical reasoning learning environment: A rasch analysis. *Advanced Science Letters*, 21(5), 1211–1215. <https://doi.org/10.1166/asl.2015.6077>
- Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2011). Facilitating high school students' use of multiple representations to describe and explain simple chemical reactions. *Teaching Science*, 57(4), 13–19.
- Gracia, S. (2005). Analyzing CSR implementation with the Rasch Model. *Faculty Publication, Rhode Island College*.
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 989–1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Hidayat, R., Hermandra, Zetriuslita, Lestari, S., & Qudratuddarsi, H. (2022). Achievement Goals, Metacognition and Horizontal Mathematization: A Mediational Analysis. *TEM Journal*, 11(4), 1537–1546. <https://doi.org/10.18421/TEM114-14>
- Hidayat, R., Idris, W. I. W., Qudratuddarsi, H., & Rahman, M. N. A. (2021). Validation of the Mathematical Modeling Attitude Scale for Malaysian Mathematics Teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(12). <https://doi.org/10.29333/EJMSTE/11375>
- Hidayat, R., Imami, M. K. W., Liu, S., Qudratuddarsi, H., & Saad, M. R. M. (2024). Validity of engagement instrument during online learning in mathematics education. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, 8(2), 398–414. <https://doi.org/10.22437/jiituj.v8i2.34453>
- Hidayat, R., Qudratuddarsi, H., Mazlan, N. H., & Mohd Zeki, M. Z. (2021). Evaluation of a test measuring mathematical modelling competency for indonesian college students. *Journal of Nusantara Studies (JONUS)*, 6(2), 133–155. <https://doi.org/10.24200/jonus.vol6iss2pp133-155>
- Hoe, K. Y., & Subramaniam, R. (2016). On the prevalence of alternative conceptions on acid-base chemistry among secondary students: Insights from cognitive and confidence measures. *Chemistry Education Research and Practice*, 17(2), 263–282. <https://doi.org/10.1039/c5rp00146c>
- Kirbulut, Z. D, & Geban, O. (2014). Using three-tier test to assess students' misconceptions of states of matter. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(5), 509–521. <https://doi.org/10.12973/eurasia.2014.1128a>
- Laliyo, L. A. R., Hamdi, S., Pikoli, M., Abdullah, R., & Panigoro, C. (2021). Implementation of Four-Tier Multiple-Choice Instruments Based on the Partial Credit Model in Evaluating Students' Learning Progress. *European Journal of Educational Research*, 10(2), 825-840. <https://doi.org/10.12973/eu-er.10.2.825>

- Lin, J. W. (2016). Development and evaluation of the diagnostic power for a computer-based two-tier assessment. *Journal of Science Education and Technology*, 25(3), 497–511. <https://doi.org/10.1007/s10956-016-9609-5>
- Liu, T. C. (2010). Developing simulation-based computer assisted learning to correct students' statistical misconceptions based on cognitive conflict theory, using correlation|| as an example. *Educational Technology and Society*, 13(2), 180–192.
- Liu, O. L., Lee, H., Linn, M. C., & Liu, O. L. (2011). An investigation of explanation multiple-choice items in science assessment. *Educational Assessment*, 16(3), 164–184. <https://doi.org/10.1080/10627197.2011.611702>
- Park, M., & Liu, X. (2019). An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics. *Research in Science Education*. <https://doi.org/10.1007/s11165-019-9819-y>
- Rappoport, L. T., & Ashkenazi, G. (2008). Connecting levels of representation: Emergent versus submergent perspective. *International Journal of Science Education*, 30(12), 1585–1603. <https://doi.org/10.1080/09500690701447405>
- Qudratuddarsi, H., Hidayat, R., Nasir, N., Imami, M. K. W., & bin Mat Nor, R. (2022). Rasch validation of instrument measuring Gen-Z science, technology, engineering, and mathematics (STEM) application in teaching during the pandemic. *International Journal of Learning, Teaching and Educational Research*, 21(6), 104-121. <https://doi.org/10.26803/ijlter.21.6.7>
- Qudratuddarsi, H., Sathasivam, R. V., & Hutkemri, H. (2019). Difficulties and Correlation between Phenomenon and Reasoning Tier of Multiple-Choice Questions: A Survey Study. *Indonesian Research Journal in Education | IRJE*, 249-264.
- Ramadhana, N., & Qudratuddarsi, H. (2024). Analisis Self Efficacy Mahasiswa pada Mata Kuliah Biologi Sel. *Saqbe: Jurnal Sains dan Pembelajarannya*, 1(1), 33-38.
- Romine, W. L., Todd, A. N., & Clark, T. B. (2016). How do undergraduate students conceptualize acid–base chemistry? Measurement of a concept progression. *Science Education*, 100(6), 1150–1183. <https://doi.org/10.1002/sce.21240>
- Santos, V. C., & Arroio, A. (2016). The representational levels: Influences and contributions to research in chemical education. *Journal of Turkish Science Education*, 13(1), 3–18. <https://doi.org/10.12973/tused.10153a>
- Schultz, M., Lawrie, G. A., Bailey, C. H., Simon, B., Dargaville, T. R., Brien, G. O., ... Thompson, C. D. (2017). Evaluation of diagnostic tools that tertiary teachers can apply to profile their students' conceptions. *International Journal of Science Education ISSN*: <https://doi.org/10.1080/09500693.2017.1296980>
- Sreenivasulu, B., & Subramaniam, R. (2014). Exploring undergraduates' understanding of transition metals chemistry with the use of cognitive and confidence measures. *Research in Science Education*, 44(6), 801–828. <https://doi.org/10.1007/s11165-014-9400-7>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Cimahi: Trim Komunikata
- Taber, Keith S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice*, 14(2), 156–168. <https://doi.org/10.1039/c3rp00012e>
- Touli, E. H., Talbi, M., & Radid, M. (2015). Barriers to learning physics in moroccan secondary college: the case of student's representations in electricity. *Procedia - Social and Behavioral Sciences*, 197(July), 278–280. 138. <https://doi.org/10.1016/j.sbspro.2015.07.136>

- Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, 89(3), 335–345. <https://doi.org/10.1021/ed100852t>
- Xiao, Y., Han, J., Koenig, K., Xiong, J., & Bao, L. (2018). Multilevel Rasch modeling of two-tier multiple choice test: A case study using Lawson's classroom test of scientific reasoning. *Physical Review Physics Education Research*, 14(2), 020104.
- Yan, Y. K., & Subramaniam, R. (2018). Using a multi-tier diagnostic test to explore the nature of students' alternative conceptions on reaction. *Chemistry Education Research and Practice*, 19, 213–226. <https://doi.org/10.1039/c7rp00143f>
- Zuidgeest, M., Hendriks, M., Koopman, L., Spreeuwenberg, P., & Rademakers, J. (2011). Comparison of a postal survey and mixed-mode survey using a questionnaire on patients' experiences with breast care. *Journal of Medical Internet Research*, 13(3). doi:10.2196/jmir.1241