Forum on Understanding and Developing Advanced Innovation for Elementary Studies in Indonesia
Volume 1, Number 1, 2026 pp. 1-14

Open Access: https://proceeding.unesa.ac.id/index.php/fundanesa

Kelayakan Instrumen Tes Kemampuan Akademik Matematika Materi Pecahan Berbasis Prosedur Newman untuk Siswa Sekolah Dasar

Rosalia Annisa Putri[1], Tri Astari[1]
[1]Universitas Jember

***Correspondence author:** triastari.pgsd@unej.ac.id

Article History
Received : 12 Juli 2026
Acccepted : 22 August 2026
Published : 01 November 2026

Keywords
instrument feasibility;
mathematics TKA;
fractions;
Newman Procedure;
elementary school;

Abstract

A mathematics Academic Ability Test intended to support the identification of students' problem-solving errors requires adequate content, construction, language, and classroom feasibility before empirical testing. This study aimed to describe the feasibility of a fraction TKA instrument based on Newman's Procedure through expert and practitioner judgments and to identify revision priorities. A descriptive evaluative design was employed at the instrument-review stage. The draft consisted of 15 items for Grade 6 elementary students, including nine multiple-choice and six complex multiple-choice items. Five validators participated: one mathematics content expert, one assessment-construction expert, one language expert, and two Grade 6 teacher practitioners. Each validation form contained ten criteria rated on a five-point scale. The obtained scores were converted into feasibility percentages. The content aspect obtained 36/50 (72%) and was categorized as feasible; construction obtained 40/50 (80%) and was categorized as feasible; language obtained 47/50 (94%) and was categorized as highly feasible; and practitioner assessment obtained 92/100 (92%) and was categorized as highly feasible. Content was identified as the main revision priority, particularly regarding alignment with indicators, problem complexity, and the ability of items to provide evidence of reading, comprehension, and transformation errors within Newman's Procedure. Therefore, the instrument is feasible to proceed to empirical testing after revision based on validator feedback.

Pendahuluan

Pecahan merupakan salah satu materi matematika yang penting sekaligus kompleks bagi siswa sekolah dasar. Pemahaman pecahan tidak hanya berkaitan dengan kemampuan melakukan operasi hitung, tetapi juga mencakup pemahaman terhadap berbagai representasi dan makna pecahan. Jiang et al. (2021) menunjukkan bahwa pemahaman siswa terhadap subkonstruk pecahan dapat berkembang secara berbeda sehingga penguasaan pada satu bentuk pecahan tidak selalu menunjukkan penguasaan pada bentuk lainnya. Oleh karena itu, instrumen asesmen pecahan perlu disusun untuk merepresentasikan kompetensi yang hendak diukur secara memadai (American Educational Research Association, 2014).

Pemahaman pecahan juga melibatkan keterkaitan antara pengetahuan konseptual dan prosedural. Penggunaan representasi dan kegiatan pembelajaran yang menghubungkan konsep dengan prosedur dapat

*This is an open access article under the **CC BY SA** license*

membantu siswa membangun pemahaman pecahan secara lebih utuh (Bruce et al., 2023; Jayanthi et al., 2021). Kondisi tersebut menunjukkan bahwa asesmen pecahan sebaiknya tidak hanya menilai kemampuan memperoleh jawaban akhir, tetapi juga memperhatikan proses siswa dalam memahami dan menyelesaikan permasalahan.

Dalam konteks kebijakan pendidikan Indonesia, Tes Kemampuan Akademik (TKA) digunakan untuk mengukur capaian akademik peserta didik pada mata pelajaran tertentu (Kementerian Pendidikan Dasar dan Menengah Republik Indonesia, 2025). Penyusunan instrumen matematika perlu memperhatikan keselarasan antara materi, kompetensi, proses kognitif, bentuk soal, dan tujuan penggunaan hasil asesmen. Hal tersebut penting karena validitas berkaitan dengan sejauh mana bukti yang tersedia mendukung interpretasi skor untuk tujuan penggunaannya (Buckley-Walker & Lipscombe, 2022).

Apabila instrumen digunakan untuk mendukung diagnosis kesalahan pemecahan masalah, informasi yang diperoleh tidak cukup hanya berupa jawaban benar atau salah. Salah satu kerangka yang dapat digunakan untuk menelusuri sumber kesalahan siswa adalah Prosedur Newman. Newman (1983) mengelompokkan proses penyelesaian masalah ke dalam lima tahap, yaitu membaca (*reading*), memahami (*comprehension*), mentransformasikan masalah (*transformation*), melakukan keterampilan proses (*process skills*), dan menuliskan jawaban akhir (*encoding*).

Penelitian pada materi pecahan menunjukkan bahwa kesalahan siswa dapat terjadi pada berbagai tahapan tersebut. Kesalahan tidak hanya muncul pada proses perhitungan, tetapi juga ketika siswa memahami informasi, menentukan operasi yang sesuai, serta menuliskan jawaban akhir (Labibah et al., 2021; Priwantoro et al., 2023; Putri & Purwanto, 2022). Oleh sebab itu, penyusunan instrumen berbasis Prosedur Newman perlu memperhatikan kemampuan setiap butir dalam menyediakan informasi yang relevan mengenai proses penyelesaian siswa.

Sebelum digunakan dalam uji coba, instrumen perlu melalui proses penelaahan untuk mengetahui kesesuaian isi, konstruk, bahasa, dan keterlaksanaannya. Penelaahan oleh ahli merupakan salah satu tahap penting dalam pengembangan instrumen sebelum dilakukan uji coba dan analisis karakteristik empiris butir (Muñiz & Fonseca-Pedrero, 2019). Selain ahli akademik, pelibatan guru sebagai praktisi dapat memberikan informasi mengenai keterpahaman dan keterlaksanaan instrumen dalam konteks pembelajaran sekolah dasar. Astari et al. (2025), misalnya, melibatkan ahli akademik dan guru sekolah dasar sebagai praktisi dalam proses penilaian instrumen untuk memperoleh pertimbangan dari sisi teoretis sekaligus keterpakaian instrumen di sekolah.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mendeskripsikan tingkat kelayakan instrumen TKA matematika materi pecahan berbasis Prosedur Newman berdasarkan penilaian aspek isi, konstruk, bahasa, dan praktisi serta mengidentifikasi aspek yang perlu diprioritaskan dalam revisi sebelum instrumen digunakan pada tahap uji coba.

Metode

Desain Penelitian

Penelitian ini menggunakan desain deskriptif evaluatif pada tahap penelaahan instrumen. Fokus penelitian adalah mendeskripsikan hasil penilaian ahli dan praktisi terhadap draf instrumen TKA matematika materi pecahan berbasis Prosedur Newman. Draf instrumen terdiri atas 15 butir soal untuk siswa kelas VI sekolah dasar, yaitu sembilan soal pilihan ganda (PG) dan enam soal pilihan ganda kompleks (PGK). Cakupan materi meliputi pecahan senilai, membandingkan dan mengurutkan pecahan, hubungan pecahan dengan desimal dan persen, serta operasi hitung pecahan. Butir soal disusun dengan mempertimbangkan karakteristik TKA dan diarahkan untuk mendukung penelusuran kesalahan berdasarkan lima tahap Prosedur Newman, yaitu membaca, memahami, transformasi, keterampilan proses, dan penulisan jawaban akhir.

Validator

Penilaian instrumen melibatkan lima validator yang dipilih berdasarkan bidang keahlian dan pengalaman. Tiga validator berasal dari kalangan dosen, masing-masing sebagai ahli isi/materi matematika sekolah dasar, ahli konstruk/asesmen, dan ahli bahasa. Dua validator lainnya merupakan guru kelas VI yang berperan sebagai praktisi.

Tabel 1. Komposisi Validator dan Fokus Penilaian

Aspek	Validator	Fokus Penilaian
Isi/Materi	1 dosen ahli matematika SD	Kesesuaian indikator, materi, kompleksitas, dan dukungan terhadap analisis Newman
Konstruk	1 dosen ahli asesmen	Struktur soal, bentuk PG/PGK, petunjuk, opsi, dan tuntutan penalaran
Bahasa	1 dosen ahli bahasa	Ejaan, keterbacaan, istilah, kejelasan perintah, dan kesesuaian bahasa dengan siswa
Praktisi	2 guru kelas VI	Relevansi materi, keterbacaan, waktu, tingkat kesulitan, administrasi, dan keterlaksanaan

Pelibatan validator dengan bidang yang berbeda dimaksudkan agar penilaian instrumen tidak hanya berfokus pada ketepatan materi, tetapi juga mencakup kualitas konstruksi, bahasa, dan keterlaksanaannya di kelas. Pendekatan yang melibatkan ahli dan praktisi juga digunakan oleh Astari et al. (2025) dalam pengembangan instrumen untuk siswa sekolah dasar.

Teknik Analisis Data

Setiap validator memberikan penilaian menggunakan skala 1–5 terhadap sepuluh kriteria yang terdapat pada lembar penilaian masing-masing. Skor kemudian dianalisis menggunakan rumus:

$$\text{Valpro} = \frac{srt}{smt} \times 100\%$$

Keterangan:

Valpro : Validitas Instrumen

srt : Skor Riil Tercapai

smt : Skor Maksimal Tercapai

Kategori kelayakan mengacu pada modifikasi kriteria Masyhud (2021).

Tabel 2. Kategori Kelayakan Instrumen

Persentase	Kategori

$80 < Valpro \leq 100$	Sangat Layak
$60 < Valpro \leq 80$	Layak
$40 < Valpro \leq 60$	Cukup Layak
$20 < Valpro \leq 40$	Kurang Layak
$0 < Valpro \leq 20$	Sangat Kurang Layak

Skor dua praktisi digabungkan karena keduanya menggunakan jumlah kriteria dan skala penilaian yang sama. Dengan demikian, skor maksimal aspek praktisi adalah 100.

Persentase yang diperoleh digunakan sebagai gambaran deskriptif kelayakan instrumen dan tidak diperlakukan sebagai *Content Validity Index (CVI)* maupun ukuran kesepakatan antarvalidator. Penghitungan CVI memerlukan desain panel dan prosedur analisis yang berbeda (Almanasreh et al., 2019; Yusoff, 2019).

Hasil dan Pembahasan

Hasil Penilaian Kelayakan Instrumen

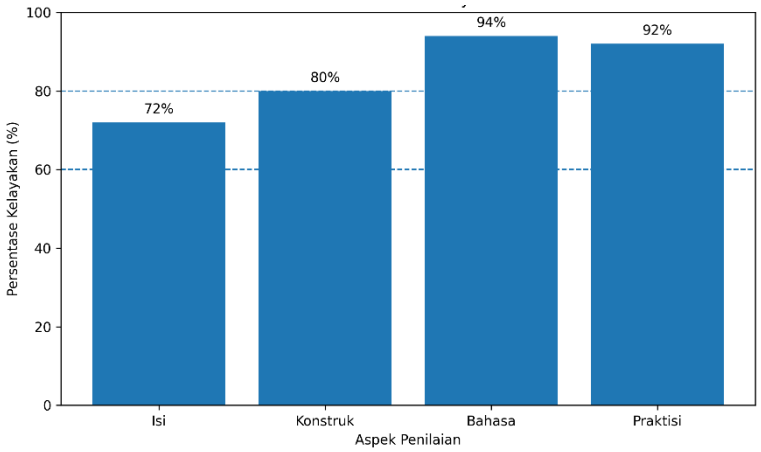
Hasil penilaian ahli dan praktisi menunjukkan adanya perbedaan tingkat kelayakan pada masing-masing aspek. Hasil analisis disajikan pada Tabel 3.

Tabel 3. Hasil Penilaian Kelayakan Instrumen

Aspek	Skor Tercapai	Skor Maksimal	Persentase	Kategori
Isi/Materi	36	50	72%	Layak
Konstruk	40	50	80%	Layak
Bahasa	47	50	94%	Sangat Layak
Praktisi	92	100	92%	Sangat Layak

Hasil penilaian menunjukkan bahwa seluruh aspek instrumen telah mencapai sekurang-kurangnya kategori layak, meskipun persentase yang diperoleh pada setiap aspek berbeda. Aspek bahasa memperoleh persentase tertinggi sebesar 94%, diikuti aspek praktisi sebesar 92%, konstruk sebesar 80%, dan isi/materi sebesar 72%. Perbedaan capaian tersebut menunjukkan bahwa kelayakan instrumen perlu ditinjau berdasarkan karakteristik masing-masing aspek dan tidak hanya didasarkan pada satu nilai gabungan. Penilaian terhadap beberapa aspek secara terpisah penting karena kualitas instrumen tidak hanya ditentukan oleh ketepatan isi, tetapi juga oleh konstruksi, keterbacaan, dan kesesuaiannya dengan konteks penggunaan. Schames Kreitchmann et al. (2024) menegaskan bahwa penilaian ahli terhadap isi instrumen perlu memperhatikan relevansi, keterwakilan, dan kesesuaian butir dengan konstruk yang hendak diukur. Untuk memperjelas perbandingan hasil penilaian pada keempat aspek tersebut, persentase kelayakan disajikan pada Gambar 1.

Gambar 1. Persentase Kelayakan Instrumen



Sumber: Data Penelitian, 2026

Berdasarkan Gambar 1, aspek bahasa dan penilaian praktisi memiliki capaian yang lebih tinggi dibandingkan aspek konstruk dan isi/materi. Aspek bahasa sebesar 94% dan praktisi sebesar 92% termasuk kategori sangat layak, sedangkan konstruk sebesar 80% dan isi/materi sebesar 72% termasuk kategori layak. Hasil tersebut menunjukkan bahwa instrumen memiliki kekuatan terutama pada penggunaan bahasa dan keterlaksanaannya dalam konteks pembelajaran kelas VI. Sementara itu, aspek isi/materi yang memperoleh persentase terendah memerlukan perhatian lebih dalam proses revisi. Hal ini sejalan dengan Buckley-Walker & Lipscombe (2022) yang menunjukkan bahwa kualitas bukti dari asesmen kelas dipengaruhi oleh keselarasan asesmen dengan kurikulum dan pembelajaran, kesesuaiannya dengan kemampuan siswa, serta bagaimana hasil asesmen akan digunakan.

Kelayakan Aspek Isi/Materi

Aspek isi/materi memperoleh skor 36 dari skor maksimal 50 atau sebesar 72% dan termasuk kategori layak. Persentase tersebut merupakan capaian terendah dibandingkan aspek lainnya sehingga menjadi prioritas dalam proses revisi. Hasil penilaian menunjukkan bahwa beberapa bagian masih perlu diperkuat, terutama keselarasan antara indikator dan butir soal, tingkat kompleksitas masalah, serta kemampuan soal dalam memberikan informasi mengenai kesalahan siswa pada tahap membaca, memahami, dan mentransformasikan masalah.

Penilaian isi memiliki kedudukan penting karena butir yang digunakan harus benar-benar merepresentasikan kemampuan yang hendak diukur. Schames Kreitchmann et al. (2024) menjelaskan bahwa ahli materi atau *subject-matter experts* berperan dalam menilai relevansi, keterwakilan, dan ketepatan butir terhadap konstruk. Dalam konteks instrumen pecahan, hal tersebut berarti bahwa soal tidak hanya harus benar secara matematis, tetapi juga harus sesuai dengan indikator dan mampu merepresentasikan aspek pengetahuan pecahan yang hendak dianalisis.

Temuan tersebut juga relevan dengan penelitian Schadl & Lindmeier (2025) yang menunjukkan bahwa pengembangan instrumen untuk menilai pembelajaran pecahan perlu didasarkan pada model kognitif yang jelas mengenai komponen pengetahuan pecahan. Instrumen yang dikembangkan untuk tujuan diagnostik perlu mampu menghasilkan informasi yang lebih spesifik mengenai pengetahuan dan kesulitan siswa, bukan hanya skor benar dan salah. Oleh karena itu, keterwakilan materi serta hubungan antara butir dan proses kognitif menjadi bagian penting dalam revisi instrumen. Dalam penelitian ini, kebutuhan revisi

isi juga berkaitan dengan penggunaan Prosedur Newman. Prosedur tersebut membedakan sumber kesalahan menjadi tahap membaca, memahami, transformasi, keterampilan proses, dan penulisan jawaban akhir (Newman, 1983). Dengan demikian, jawaban siswa yang salah belum tentu menunjukkan satu jenis kesalahan tertentu. Kesalahan dapat terjadi karena siswa tidak memahami informasi, salah menentukan operasi, melakukan kesalahan perhitungan, atau tidak tepat menuliskan jawaban akhir.

Temuan penelitian terbaru pada materi pecahan memperkuat pertimbangan tersebut. Labibah et al. (2021) menemukan adanya kesalahan siswa pada beberapa tahapan Newman ketika menyelesaikan soal cerita pecahan. Putri & Purwanto (2022) juga menunjukkan bahwa siswa sekolah dasar dapat mengalami kesalahan setelah tahap membaca, terutama dalam memahami, mentransformasikan, melakukan proses penyelesaian, dan menuliskan jawaban akhir. Sementara itu, Priwantoro et al. (2023) menemukan kesalahan pada pemahaman konsep pecahan, transformasi representasi, proses perhitungan, dan penyajian jawaban. Temuan-temuan tersebut menunjukkan bahwa identifikasi kesalahan pecahan memerlukan butir yang mampu menyediakan informasi mengenai proses penyelesaian siswa.

Berdasarkan hasil tersebut, revisi aspek isi diarahkan pada penajaman hubungan antara indikator, materi pecahan, dan tahap Newman yang akan dianalisis. Pilihan jawaban atau pengecoh pada soal PG dan PGK juga perlu dikembangkan berdasarkan kemungkinan pola kesalahan siswa sehingga tidak hanya berfungsi sebagai alternatif jawaban, tetapi dapat memberikan informasi awal mengenai jenis kesalahan yang mungkin dilakukan siswa. Namun, penetapan jenis kesalahan tetap perlu didukung oleh bukti proses penyelesaian siswa agar interpretasi tidak hanya didasarkan pada pilihan jawaban.

Kelayakan Aspek Konstruk

Aspek konstruk memperoleh skor 40 dari 50 atau 80% dan termasuk kategori layak. Berdasarkan kategori yang digunakan dalam penelitian ini, nilai 80% belum termasuk kategori sangat layak karena kategori sangat layak berlaku pada nilai lebih dari 80%. Hasil tersebut menunjukkan bahwa bentuk dan struktur instrumen secara umum telah memenuhi kriteria kelayakan, meskipun masih diperlukan beberapa penyempurnaan.

Aspek konstruk mencakup penggunaan bentuk soal PG dan PGK, kejelasan petunjuk, susunan pertanyaan, hubungan antara pokok soal dan pilihan jawaban, serta tuntutan penalaran. Perbaikan terutama diperlukan pada kejelasan petunjuk PGK dan hubungan antara pertanyaan dengan pilihan jawaban. Petunjuk yang kurang jelas dapat menyebabkan respons siswa dipengaruhi oleh ketidakpahaman terhadap mekanisme pengerjaan, bukan oleh kemampuan matematika yang sedang diukur. Buckley-Walker & Lipscombe (2022) menunjukkan bahwa dalam desain asesmen matematika di sekolah dasar, keselarasan antara kurikulum, pembelajaran, karakteristik siswa, serta prosedur pemberian skor merupakan unsur yang berhubungan dengan kualitas bukti asesmen. Dengan demikian, konstruksi butir perlu memastikan bahwa tuntutan soal sejalan dengan kompetensi yang hendak diukur dan tidak menimbulkan sumber kesulitan yang tidak berkaitan dengan materi.

Selain itu, Schadl & Lindmeier (2025) menekankan bahwa perubahan atau pengembangan bentuk suatu tes tetap perlu diikuti dengan pemeriksaan kualitas instrumen sesuai tujuan penggunaannya. Hal ini mendukung perlunya penyempurnaan bentuk PG dan PGK dalam penelitian ini sebelum instrumen memasuki tahap uji coba empiris. Dengan kata lain, hasil penilaian konstruk sebesar 80% menunjukkan bahwa instrumen telah memiliki dasar struktur yang memadai, tetapi masih perlu dioptimalkan agar respons siswa lebih merepresentasikan kemampuan pecahan yang dimaksud.

Kelayakan Aspek Bahasa

Aspek bahasa memperoleh skor 47 dari 50 atau 94% dan termasuk kategori sangat layak. Persentase tersebut merupakan capaian tertinggi dari seluruh aspek yang dinilai. Hasil ini menunjukkan bahwa bahasa dalam instrumen secara umum telah memenuhi kriteria keterbacaan, ketepatan penggunaan istilah matematika, struktur kalimat, kejelasan perintah, dan kesesuaian bahasa dengan karakteristik siswa kelas VI. Meskipun memperoleh persentase yang tinggi, penilaian ahli bahasa belum sepenuhnya membuktikan bahwa setiap siswa akan memahami stimulus dan perintah dengan cara yang dimaksudkan oleh penyusun instrumen. Bahasa dalam asesmen perlu dipastikan tidak justru menjadi sumber kesulitan tambahan yang tidak berkaitan dengan konstruk matematika. Kūkea Shultz & Englert (2021) menunjukkan bahwa pemeriksaan proses respons, termasuk melalui wawancara kognitif, dapat memberikan informasi mengenai bagaimana siswa memahami kosakata, bahasa, dan maksud suatu butir ketika mengerjakan asesmen.

Pertimbangan tersebut menjadi semakin penting pada instrumen berbasis Prosedur Newman karena membaca dan memahami merupakan dua tahap awal yang dianalisis. Apabila kalimat soal atau istilah yang digunakan tidak dipahami siswa, peneliti perlu membedakan apakah kesulitan tersebut merupakan karakteristik kemampuan siswa yang memang ingin didiagnosis atau justru disebabkan oleh redaksi butir yang kurang tepat. Oleh karena itu, meskipun aspek bahasa memperoleh kategori sangat layak, uji coba terbatas kepada siswa tetap diperlukan untuk memastikan bahwa stimulus, istilah, dan petunjuk soal dipahami sebagaimana yang dimaksudkan.

Kelayakan Berdasarkan Penilaian Praktisi

Penilaian praktisi memperoleh skor gabungan 92 dari 100 atau 92% dan termasuk kategori sangat layak. Hasil tersebut menunjukkan bahwa instrumen dipandang dapat diterapkan pada siswa kelas VI ditinjau dari relevansi materi, keterbacaan, waktu pengerjaan, tingkat kesulitan, stimulus, tata letak, dan kemudahan pelaksanaannya. Pelibatan guru sebagai praktisi memberikan perspektif yang berbeda tetapi melengkapi penilaian dosen ahli. Guru memiliki pengalaman langsung mengenai karakteristik siswa, situasi pembelajaran, kecukupan waktu, dan kemungkinan pelaksanaan asesmen di kelas. Astari et al. (2025) juga melibatkan akademisi dan guru sekolah dasar sebagai praktisi dalam validasi instrumen dan menunjukkan bahwa keterlibatan kedua kelompok tersebut dapat memberikan pertimbangan dari sisi akademik-teoretis sekaligus praktis-kontekstual.

Hal ini juga sejalan dengan Buckley-Walker & Lipscombe (2022) yang menemukan bahwa guru dalam merancang asesmen kelas mempertimbangkan keselarasan dengan pembelajaran, kemampuan siswa, prosedur penilaian, serta pemanfaatan informasi yang diperoleh dari asesmen. Oleh karena itu, persentase praktisi sebesar 92% memberikan dukungan bahwa instrumen secara umum sesuai dengan kondisi penggunaan di kelas VI. Meskipun demikian, skor praktisi tersebut merupakan persentase kelayakan dan bukan ukuran reliabilitas atau kesepakatan antarpenilai. Dua praktisi memberikan penilaian terhadap kriteria yang sama, kemudian skornya digabungkan sesuai prosedur analisis penelitian. Oleh karena itu, hasil ini lebih tepat ditafsirkan sebagai gambaran kelayakan instrumen dari perspektif pengguna di lapangan.

Implikasi Hasil Penilaian

Secara keseluruhan, hasil penilaian menunjukkan bahwa instrumen memiliki kelayakan yang memadai untuk dilanjutkan ke tahap uji coba setelah dilakukan revisi. Aspek bahasa dan keterlaksanaan menurut praktisi merupakan bagian dengan capaian tertinggi, sedangkan isi/materi menjadi aspek yang paling memerlukan perbaikan. Hasil tersebut sekaligus menunjukkan bahwa penilaian ahli dan praktisi merupakan tahap awal dalam pengembangan instrumen dan belum cukup untuk menyatakan bahwa instrumen telah memiliki kualitas pengukuran secara menyeluruh. Schames Kreitchmann et al. (2024) menempatkan penilaian ahli sebagai sumber penting untuk memeriksa kualitas isi, sedangkan kualitas

instrumen selanjutnya tetap perlu ditinjau menggunakan bukti empiris. Pada konteks asesmen pecahan, Schadl & Lindmeier (2025) juga menunjukkan bahwa instrumen yang dikembangkan berdasarkan model kognitif tetap memerlukan pemeriksaan terhadap sifat psikometrik dan kesesuaiannya dengan tujuan penggunaan.

Oleh karena itu, formulasi yang lebih tepat untuk penelitian ini adalah bahwa instrumen TKA matematika materi pecahan berbasis Prosedur Newman dinilai layak berdasarkan penilaian ahli dan praktisi serta dapat dilanjutkan ke tahap uji coba setelah dilakukan revisi. Tahap uji coba selanjutnya diperlukan untuk memperoleh bukti mengenai fungsi butir pada siswa, tingkat kesukaran, daya pembeda, konsistensi hasil pengukuran, dan kecukupan bukti proses respons sebelum instrumen digunakan untuk menarik kesimpulan diagnostik yang lebih luas.

Simpulan

Instrumen TKA matematika materi pecahan berbasis Prosedur Newman menunjukkan tingkat kelayakan yang berbeda pada setiap aspek penilaian. Aspek isi/materi memperoleh persentase sebesar 72% dengan kategori layak, aspek konstruk sebesar 80% dengan kategori layak, aspek bahasa sebesar 94% dengan kategori sangat layak, dan penilaian praktisi sebesar 92% dengan kategori sangat layak. Capaian tersebut menunjukkan bahwa kekuatan instrumen terletak pada aspek bahasa dan keterlaksanaannya dalam konteks pembelajaran kelas VI, sedangkan aspek isi/materi masih memerlukan perhatian lebih dalam proses revisi. Perbaikan terutama diarahkan pada keselarasan antara indikator dan butir soal, tingkat kompleksitas masalah, serta kemampuan butir dalam memberikan informasi mengenai kesalahan membaca, memahami, dan mentransformasikan masalah berdasarkan Prosedur Newman. Berdasarkan hasil penilaian ahli dan praktisi, instrumen telah memenuhi kriteria kelayakan untuk dilanjutkan ke tahap uji coba setelah dilakukan revisi sesuai masukan validator. Uji coba empiris tetap diperlukan untuk memperoleh informasi lebih lanjut mengenai kualitas butir, karakteristik respons siswa, dan konsistensi hasil pengukuran sebelum instrumen digunakan dalam cakupan yang lebih luas.

Daftar Pustaka

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

American Educational Research Association, A. P. A. & N. C. on M. in E. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Astari, T., Nurhasanah, Cahyanita, E., & Nuriman. (2025). Pengembangan instrumen literasi sains berbasis kearifan lokal untuk pembelajaran di sekolah dasar. *Prosiding Seminar Nasional Pendidikan Sains (SNPS) 2025. Prosiding Universitas Sebelas Maret*.

Bruce, C. D., Flynn, T., Yearley, S., & Hawes, Z. (2023). Leveraging Number Lines and Unit Fractions to Build Student Understanding: Insights from a Mixed Methods Study. *Canadian Journal of Science, Mathematics and Technology Education*, *23*(2), 322–339. https://doi.org/10.1007/s42330-023-00278-x

Buckley-Walker, K., & Lipscombe, K. (2022). Validity and the design of classroom assessment in teacher teams. *The Australian Educational Researcher*, *49*(2), 425–444. https://doi.org/10.1007/s13384-021-00437-9

Jayanthi, M., Gersten, R., Schumacher, R. F., Dimino, J., Smolkowski, K., & Spallone, S. (2021). Improving Struggling Fifth-Grade Students' Understanding of Fractions: A Randomized Controlled Trial of an Intervention That Stresses Both Concepts and Procedures. *Exceptional Children*, *88*(1), 81–100. https://doi.org/10.1177/00144029211008851

Jiang, Z., Mok, I. A. C., & Li, J. (2021). Chinese Students' Hierarchical Understanding of Part-whole and Measure Subconstructs. *International Journal of Science and Mathematics Education*, *19*(7), 1441–1461. https://doi.org/10.1007/s10763-020-10118-1

Kementerian Pendidikan Dasar dan Menengah Republik Indonesia. (2025). *Peraturan Menteri Pendidikan Dasar dan Menengah Republik Indonesia Nomor 9 Tahun 2025 tentang Tes Kemampuan Akademik*. Https://Peraturan.Bpk.Go.Id/Details/321171/Permendikdasmen-No-9-Tahun-2025.

Kūkea Shultz, P., & Englert, K. (2021). Cultural Validity as Foundational to Assessment Development: An Indigenous Example. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.701973

Labibah, N., Damayani, A. T., & Sary, R. M. (2021). Analisis Kesalahan Siswa Berdasarkan Teori Newman dalam Menyelesaikan Soal Cerita pada Materi Pecahan Kelas V Madrasah Ibtidaiyah. *Journal for Lesson and Learning Studies*, *4*(2), 208–216. https://doi.org/10.23887/jlls.v4i2.33265

Masyhud, M. S. (2021). *Metode penelitian pendidikan: Penuntun teori dan praktik penelitian bagi calon guru, guru, dan praktisi pendidikan (Edisi ke-7)*. Lembaga Pengembangan Manajemen dan Profesi Kependidikan.

Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, *1*(31), 7–16. https://doi.org/10.7334/psicothema2018.291

Newman, A. (1983). *The Newman language of mathematics kit: Strategies for diagnosis and remediation*. Harcourt Brace Jovanovich.

Priwantoro, S. W., Monica, V., Fauzia, N. A., & Larasati, L. (2023). Common errors analysis in solving problems of fractional number reviewed from Newman's theory. *International Journal on Education Insight*, *4*(2), 65–72. https://doi.org/10.12928/ijei.v4i2.10291

Putri, M. A., & Purwanto, S. E. (2022). Analisis Kesalahan Siswa SD Kelas V dalam Menyelesaikan Masalah Soal Cerita pada Materi Pecahan Berdasarkan Prosedur Newman. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *7*(1), 1–15. https://doi.org/10.31004/cendekia.v7i1.1653

Schadl, C., & Lindmeier, A. (2025). Searching for efficient and informative instruments to digitally monitor fraction learning: putting tests for fraction subconstruct knowledge and informal fraction knowledge on the test bench. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1367942

Schames Kreitchmann, R., Nájera, P., Sanz, S., & Sorrel, M. Á. (2024). Enhancing content validity assessment with item response theory modeling. *Psicothema*, *36*(2), 145–153. https://doi.org/10.7334/psicothema2023.208

Yusoff, M. S. B. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal*, *11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6