



## TWO-TIERS TEST BERBASIS EVERYDAY LIFE ACTIVITY STIMULUS UNTUK MENINGKATKAN KEMAMPUAN NUMERASI AKM SISWA SEKOLAH DASAR

Wilis Putri Hapsari<sup>1</sup> & Umar Abdul Labib<sup>2</sup>

<sup>1</sup>Sekolah Dasar Negeri Bojong Baru, Kulon Progo, Yogyakarta, Indonesia

<sup>2</sup>Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

<sup>1</sup>Contributor Email: [wilishapsari7@gmail.com](mailto:wilishapsari7@gmail.com)

Received: June 18, 2024

Accepted: October 30, 2024

Published: March 30, 2025

Article Url: <https://ojsdikdas.kemendikdasmen.go.id/index.php/didaktika/article/view/1764>

### Abstract

*This research investigated the effectiveness of numeracy assessment with two-tier model and daily life activity stimulus instruments in improving numeracy outcomes. The numeracy instruments were administered to 442 elementary school students in Kulon Progo Regency using random sampling techniques. Instrument validity was established using Aiken's formula for content validity and Exploratory Factor Analysis (EFA) for construct validity. Reliability was determined using Cronbach's Alpha. Test quality was analyzed using Item Response Theory (IRT) through Partial Credit Model (PCM) and descriptive analysis to compare the differences in AKM results between 2022 and 2023. The findings indicated that the instrument met the content validity criteria with an Aiken's index exceeding 0.7 and a KMO-MSA value of 0.78 for construct validity, along with a significant Bartlett's value of 0.00, confirming construct validity. Reliability was categorized as highly reliable with a Cronbach's Alpha of 0.872. The instrument could measure student ability within a range of -1.2 to 2.8, providing a maximum information function of 58 and a standard error of 0.7. Numeracy achievement showed an increase of 16.01% compared from years before (30.60%) to the years after (46.67%) the test was held. Therefore, it can be concluded that the two-tier based daily life activity stimulus numeracy assessment instrument possesses good quality and can effectively enhance numeracy achievement.*

**Keywords:** *Assesment Instrument; Numeracy; Elementary School.*

---

## Abstrak

Penelitian ini bertujuan untuk membuktikan kualitas instrumen penilaian numerasi two-tier dengan stimulus aktivitas sehari-hari dalam meningkatkan hasil capaian numerasi. Instrumen numerasi diujikan pada 442 siswa Sekolah Dasar (SD) di Kabupaten Kulon Progo yang diambil dengan teknik random sampling. Validitas instrumen dibuktikan dengan formula Aiken untuk validitas isi dan Exploratory Factor Analysis untuk validitas konstruk. Reliabilitas instrumen dibuktikan dengan Cronbach's Alpha. Kualitas tes dianalisis menggunakan Item Response Theory melalui uji Partial Credit Model dan analisis deskriptif untuk membandingkan perbedaan hasil AKM tahun 2022 dan tahun 2023. Hasil penelitian menunjukkan bahwa instrumen yang digunakan memenuhi kriteria validitas isi dengan indeks Aiken lebih dari 0,7 dan nilai KMO-MSA pada validitas konstruksinya sebesar 0,78 dengan nilai signifikansi Bartlett's 0,00 yang berarti memenuhi validitas konstruk. Reliabilitas berada pada kategori sangat andal dengan nilai Cronbach's Alpha sebesar 0,872. Instrumen yang digunakan dapat mengukur kemampuan siswa pada rentang -1,2 sampai 2,8 dengan memberikan fungsi informasi maksimal sebesar 58 dengan standar error 0,7. Capaian kemampuan numerasi terbukti mengalami kenaikan sebanyak 16,01% dibandingkan hasil capaian numerasi sebelum (30,60%) dan sesudah tes (46,67%) dilakukan. Dapat disimpulkan bahwa instrumen penilaian numerasi two-tier test berbasis aktivitas sehari-hari layak digunakan untuk meningkatkan capaian kemampuan numerasi.

**Kata Kunci:** Instrumen Asesmen; Numerasi; Sekolah Dasar.

---

## A. Pendahuluan

Peserta didik diharapkan dapat memperoleh pengalaman belajar yang optimum melalui pembelajaran dalam berbagai macam kondisi. Di samping pembelajaran inti seperti membaca, menulis, dan menghitung pada masa ini peserta didik diharuskan juga mampu mengoptimalkan kemampuan mereka untuk terlibat dalam solusi dari permasalahan global yang di antaranya berhubungan dengan kesadaran sosial, finansial, kesehatan, dan permasalahan di lingkungan sekitar (Cobo, 2013). Kemampuan anak dalam literasi dan numerasi lebih dapat diwariskan daripada intelegensi yang dimiliki oleh orang tuanya (Kovas et al., 2013). Hal tersebut memberikan gambaran kemampuan literasi dan numerasi siswa yang disesuaikan dengan lingkungan belajar rumah, kemampuan orang tua, ataupun fasilitas yang disediakannya. Konteks pembelajaran juga harus menyesuaikan dinamika sektor kehidupan manusia yang selanjutnya diterapkan dalam kebijakan dan manajemen pendidikan di semua jenjang (Khasanah et al., 2020). Penyesuaian tersebut dapat berupa inovasi yang perlu dihadirkan untuk menjawab dinamika kebutuhan manusia.

Salah satu bentuk inovasi dalam sistem pendidikan nasional yang dewasa ini dikembangkan adalah sistem penilaian berbasis komputer yang disebut sebagai Asesmen Nasional Berbasis Komputer (ANBK) dengan proses Asesmen Kompetensi Minimum (AKM), survei karakter, dan survei lingkungan belajar yang pertama kali dilaksanakan pada bulan November tahun 2021 lalu berdasarkan publikasi resmi dari pemerintah (Kemendikbud, 2020). AKM tersebut dilaksanakan dengan model Multistage Adaptive Test (MSAT) yang berarti pertanyaan yang dikerjakan berdasarkan kemampuan responden. MSAT merupakan salah satu jenis dari kelompok Computer Adaptive Test (CAT) di mana sifat adaptabilitas program menuntut adanya sistem inferensi untuk menentukan keputusan bahwa setiap siswa harus menerima jumlah butir tes yang sesuai dengan karakteristik kemampuannya (Haryanto, 2013; Zanabazar et al., 2023).

Dilansir dari website resmi penyelenggaraan ANBK <https://anbk.kemdikbud.go.id> (September 2021), tampilan output capaian hasil belajar AKM tahun 2021 di wilayah Kabupaten Kulon Progo yang mencakup 354 satuan pendidikan dapat dilihat pada Gambar 1. Hasil tersebut menunjukkan bahwa capaian hasil belajar kategori kemampuan numerasi berada pada zona kuning yang masih di bawah kompetensi minimum. Sedangkan kemampuan literasi telah mencapai kompetensi minimum dengan sebagian besar siswa telah mencapai batas kompetensi minimum untuk literasi membaca namun perlu upaya mendorong lebih banyak siswa menjadi mahir. Indeks karakter di wilayah Kabupaten Kulon Progo juga sudah berkembang dengan siswa terbiasa menerapkan nilai-nilai karakter pelajar Pancasila dalam kehidupan sehari-hari.

Capaian Hasil Belajar		
<b>Kemampuan Literasi</b>	<b>Kemampuan Numerasi</b>	<b>Indeks Karakter</b>
<b>Mencapai kompetensi minimum</b>	<b>Di bawah kompetensi minimum</b>	<b>Berkembang</b>
Sebagian besar siswa telah mencapai batas kompetensi minimum untuk literasi membaca namun perlu upaya mendorong lebih banyak siswa menjadi mahir	Kurang dari 50% siswa telah mencapai batas kompetensi minimum untuk numerasi	Siswa terbiasa menerapkan nilai-nilai karakter pelajar pancasila yang berakhlak mulia, bergotong royong, mandiri, kreatif dan bernalar kritis serta berkebinekaan global dalam kehidupan sehari-hari.

*Gambar 1. Capaian Numerasi*

Rendahnya kemampuan numerasi tersebut salah satunya karena selama ini siswa Indonesia tidak terbiasa dengan soal yang berhubungan dengan pemodelan matematis, serta permasalahan mengenai keberadaan buku teks yang kurang memfasilitasi siswa untuk menerjemahkan bentuk matematika ke dalam masalah sehari-hari (Jupri et al., 2014; Munayati et al., 2015; Zahroh et al., 2020; Putri et al., 2022).

Miskonsepsi matematika yang dipersepsikan oleh guru-guru juga mendukung kurangnya kemampuan siswa dalam menyelesaikan permasalahan numerasi. Salah satu miskonsepsi yang ditemukan di lapangan berdasarkan wawancara dengan guru di antaranya adalah pengaplikasian konsep numerasi. Numerasi hanya dianggap sebagai kegiatan menghitung dan hanya dapat disampaikan oleh guru matematika atau pun hanya sewaktu pelajaran matematika. Oleh karena itu numerasi sering dipersepsikan sebagai matematika. Hal tersebut mengakibatkan kebanyakan siswa malas dan tidak menyukai pembelajaran matematika. Numerasi seharusnya tidak hanya mencakup pelajaran matematika, tetapi terlibat dalam setiap mata pelajaran (Bennison et al., 2020; Khoiriah, 2022). Guru yang paham numerasi akan melibatkan penerapan pemahaman matematis dan keterampilan pemecahan masalah untuk menyelesaikan permasalahan sehari-hari (Dole & Geiger, 2020; Azhari et al., 2024). Pembelajaran berbasis numerasi tersebut diharapkan dapat dipraktikkan guru dalam pelaksanaan kegiatan belajar sehari-hari.

Pembelajaran yang berorientasi terhadap numerasi akan memperkaya pengalaman peserta didik dalam menyelesaikan permasalahannya dalam kehidupan sehari-hari (Geiger et al., 2015; Zanabazar et al., 2023). Konteks aktivitas kehidupan sehari-hari atau *everyday life activity* sangat penting untuk memvisualisasikan masalah (Rennie & Parker, 1996). Hal ini sejalan dengan pemikiran Asrizal et al. (2018) yang menyatakan bahwa pendidikan di sekolah perlu menghasilkan lulusan yang tidak hanya memiliki pengetahuan yang relevan, tetapi juga salah satunya adalah keterampilan literasi dalam kehidupan sehari-hari mereka.

Oleh karena itu, guru harus mampu mengaitkan konsep matematis terhadap kebutuhan kehidupan sehari-hari yang tidak lepas dengan mata

pelajaran yang lainnya. Guru yang berpengalaman dalam mengajarkan literasi dan numerasi setidaknya memiliki pengetahuan dan pemahaman serta strategi dan keterampilan dalam mengintegrasikan konsep numerasi dalam pembelajaran (Md-Ali et al., 2016). Kenyataannya, masih terdapat pendidik yang kesulitan sehingga kurang dapat mengintegrasikan konsep matematis tersebut dalam kebutuhan hidup sehari-hari. Sulitnya pendidik mengintegrasikan konsep matematis ke dalam mata pelajaran lainnya ataupun dalam penerapan kehidupan sehari-hari menjadikan soal matematika yang dibuat oleh pendidik kurang aplikatif. Berdasarkan observasi lapangan selama ini guru hanya memberikan soal matematika berupa soal cerita kepada peserta didik dalam rangka persiapan AKM. Terlebih lagi, guru juga masih kesulitan dalam menerapkan model numerasi dalam pembelajaran yang menggunakan alat terutama teknologi digital (Goos et al., 2011; Putri et al., 2022). Hal tersebut mengakibatkan peserta didik kurang mendapatkan pengalaman dalam mengerjakan soal-soal berbasis komputer ataupun pembelajaran yang terintegrasi dengan teknologi.

Instrumen penilaian mempunyai pengaruh yang besar terhadap kemampuan siswa dalam memecahkan masalah (Phillips & Phillips, 2016). Apabila siswa terbiasa mengerjakan tes yang sejalan dengan kemampuan yang akan diukur maka akan menunjukkan hasil yang baik. Siswa juga akan semakin terlatih dalam menyelesaikan masalah. Semakin terbiasanya peserta didik melatih akurasi dan efisiensi dalam mengerjakan soal akan semakin mempertajam kemampuan matematis (Kilpatrick et al., 2001; Putri et al., 2022). Instrumen tes efektif meningkatkan pemahaman, terutama pada tes yang menuntut siswa untuk menyelesaikan permasalahan di dunia nyata (Achmetli et al., 2019). Oleh karena itu kemampuan peserta didik dapat dilatih melalui pengerjaan tes (Aziza, 2019; Hamidy & Jailani, 2019). Keberadaan instrumen yang mampu mengintegrasikan konsep matematis dalam kehidupan sehari-hari masih terbatas. Butir soal yang dilengkapi dengan konteks stimulus yang berkesinambungan seputar permasalahan sehari-hari perlu dihadirkan. Hal tersebut bermaksud untuk menstimulasi pemahaman matematis dalam kehidupan sehari-hari agar peserta didik sadar akan pentingnya kebutuhan

matematika yang selama ini dipelajari. Instrumen yang dibutuhkan juga perlu menyediakan berbagai macam bentuk soal yang akan memberikan pengalaman bagi peserta didik untuk menyelesaikan permasalahan dengan berbagai macam cara.

Instrumen yang dikembangkan perlu memuat konstruksi yang digunakan dalam AKM dimana harus memperhatikan tingkat kognitif. Tingkat kognitif menunjukkan proses berpikir yang dituntut atau diperlukan untuk dapat menyelesaikan masalah atau soal. Proses kognitif pada literasi matematika-numerasi adalah *knowing* (pengetahuan dan pemahaman), *applying* (penerapan), dan *reasoning* (penalaran). Sedangkan konteks menunjukkan aspek kehidupan atau situasi untuk konten yang digunakan. Konteks pada AKM dibedakan menjadi tiga, yaitu personal, sosial budaya, dan saintifik. Konteks tersebut akan lebih efektif apabila dikontekskan pada bahasan-bahasan yang memiliki aspek kedekatan pada keseharian peserta didik.

Soal pada AKM disajikan dengan berbasis stimulus yang merangsang kemampuan berpikir anak dalam menyelesaikan masalah tertentu. Selaras dengan pendapat Verschaffel et al. (2020) yang menyatakan bahwa soal cerita merupakan permasalahan yang sulit bagi siswa dalam belajar matematika, tipe soal AKM justru menyajikan hal tersebut untuk menambah pengalaman belajar siswa.

Siswa perlu sering dihadapkan dengan permasalahan yang dekat dengan lingkungannya agar lebih kontekstual dan aplikatif dalam proses berpikir siswa yang memiliki refleksinya sendiri dan dapat membuat kesimpulan berdasarkan pengalamannya. Setiap siswa memiliki pemicu pembelajaran dengan cara yang berbeda. Hal tersebut berarti seseorang dapat menerima suatu dorongan belajar yang sama dan meresponsnya dengan cara yang berbeda (Philip et al., 2016; Shahbari et al., 2020; Ziaurrahman et al., 2024). Sehingga model soal dengan stimulus akan menjadi pemicu yang baik untuk menghasilkan proses berpikir yang sistematis. Siswa juga membutuhkan soal kontekstual yang memfasilitasi mereka untuk membuktikan proses elaborasi pemahaman terhadap cara penyelesaian masalahnya pada soal tersebut (Clarke, 2022; Li et al., 2021; Yang et al., 2015). Model tes yang sesuai dengan kebutuhan tersebut adalah model tes dua tingkat (*two-tier test*) yang dapat

menyajikan alasan siswa dalam menjawab persoalan yang sedang dikerjakannya (Halim et al., 2018; Muzaffar et al., 2020). Tes dua tingkat juga dapat mengurangi kebiasaan siswa yang menjawab tanpa dasar atau asal memilih jawaban, skor juga diperhitungkan apabila siswa mampu menjawab dengan benar pada tingkat pertama ataupun tingkat kedua.

Ketidaktercapaiannya kompetensi minimum numerasi di wilayah Kabupaten Kulon Progo mengindikasikan proses pembelajaran yang membutuhkan kontekstualisasi pembelajaran yang bermakna termasuk pada sistem penilaian yang lebih spesifik kepada instrumen penilaian yang dapat digunakan sebagai pembiasaan dorongan untuk menyelesaikan permasalahan numerasi sehingga dapat memiliki kompetensi yang layak. *Framework* AKM dapat dijadikan sumber acuan karena sudah memiliki berbagai bentuk soal dan kontekstualisasi stimulus yang memperkaya pengalaman belajar siswa. Keterkaitan konteks stimulus dengan kehidupan di lingkungan sekitar siswa juga dapat dijadikan sebagai stimulasi untuk mengupayakan pembelajaran yang lebih berorientasi pada literasi numerasi. Di samping itu juga dibutuhkan instrumen tes yang dapat memfasilitasi siswa untuk mengelaborasi pemahaman penyelesaian soalnya.

Berdasarkan diskursus tersebut peneliti tertarik untuk mengembangkan instrumen asesmen numerasi dengan *framework* AKM didukung kontekstualisasi stimulus berbasis karakter yang dekat dengan siswa bertipe soal dua tingkat untuk siswa sekolah dasar di wilayah Kabupaten Kulon Progo.

## **B. Metode**

Jenis penelitian ini adalah penelitian pengembangan (*Research and Development*) yang bertujuan mengembangkan produk berupa instrumen asesmen numerasi berbasis AKM (Asesmen Kompetensi Minimum) untuk siswa pada jenjang sekolah dasar. Produk tersebut memuat instrumen tes AKM yang disajikan berbasis website dan rubrik penskoran. Instrumen tes ini diteliti kualitas dari segi validitas dan reliabilitasnya oleh para ahli melalui *expert judgement* dan hasil tes dianalisis menggunakan teori respon butir.

Prosedur pengembangan tes yang dilakukan adalah modifikasi langkah pengembangan instrumen yang dikemukakan oleh Mardapi (2007) dan

prosedur pengembangan instrumen yang dikemukakan oleh Retnawati (2016). Kedua langkah pengembangan tersebut dimodifikasi dengan pertimbangan untuk menghasilkan tahapan yang lebih praktis digunakan, sehingga menemukan prosedur pengembangan yang terdiri atas: (1) perancangan tes yang meliputi: (a) penentuan tujuan tes, (b) penentuan kompetensi, (c) penentuan konstruk materi, (d) penyusunan kisi-kisi instrumen tes, (e) penyusunan item tes; (2) validitas isi; (3) uji coba instrumen ke-1; (4) analisis validitas konstruk dan reliabilitas; (5) perbaikan instrumen; (6) uji coba instrumen ke-2; (7) analisis dan interpretasi hasil ketercapaian; dan (8) perakitan tes.

Penelitian pengembangan ini menguji subyek penelitian menjadi dua yaitu subjek uji coba instrumen pertama dan subjek uji coba instrumen kedua yang akan diukur kemampuan numerasinya. Subjek penelitian uji coba pertama tersebut adalah siswa kelas V sekolah dasar di wilayah Kabupaten Kulon Progo. Sedangkan subjek uji coba kedua yang akan diukur kemampuan numerasinya adalah siswa kelas V lainnya di wilayah Kabupaten Kulon Progo yang tidak menjadi subyek uji coba pertama. Teknik sampling yang digunakan adalah random sampling. Teknik sampling tersebut digunakan dengan melakukan penyebaran link instrumen uji coba kepada pengawas sekolah yang mewakili tiap Kapanewon di wilayah Kabupaten Kulon Progo. Responden sebagai sampel penelitian tidak dibatasi pada kriteria tertentu ataupun klaster tertentu sehingga bersifat random. Jumlah sampel subjek uji coba pertama yang diukur adalah 200 siswa, sedangkan jumlah subjek uji coba kedua adalah 242 siswa.

Analisis validitas isi menganalisis hasil evaluasi yang diperoleh dari validator. Validitas isi dibagi menjadi dua, yakni validitas tampak dan validitas logis (Allen & Yen, 1979; Walidin et al., 2015). Validitas logis merupakan validitas yang memfokuskan apakah suatu instrumen dapat mewakili ciri-ciri indikator yang akan diukur. Proses penilaian terhadap validitas logis tersebut dilakukan untuk membuktikan instrumen AKM numerasi yang dikembangkan apakah telah mewakili konstruk instrumen yang dimaksud. Penelitian ini menggunakan formula yang dikemukakan oleh Aiken (1980). Validitas konstruk menggunakan analisis faktor eksplorasi atau

Exploratory Factor Analysis (EFA). Kriteria yang digunakan sebagai dasar untuk memeriksa efektivitas adalah item dengan beban elemen lebih besar dari 0,5 yang juga akan dipertahankan sebagai item yang valid. Jika item memiliki muatan faktor lebih besar dari satu yang mendekati satu di atas 0,3, maka item tersebut akan dibuang karena tidak satu dimensi. Analisis EFA menggunakan bantuan aplikasi SPSS. Analisis reliabilitas instrumen pada pengembangan instrumen ini menggunakan cronbach alpha.

Kualitas item soal dianalisis menggunakan teori respon butir atau *Item Response Theory* (IRT). Analisis tersebut dimaksudkan untuk mengetahui kemampuan responden tanpa dipengaruhi oleh karakteristik butir yang dikerjakannya. Namun sebelumnya dilakukan uji asumsi teori respon butir yang terdiri atas asumsi unidimensi, independensi lokal, dan invariansi parameter. Setelah uji asumsi dilakukan, analisis butir instrumen numerasi yang memiliki data politomus sehingga dianalisis dengan IRT model yakni *Partial Credit Model* (PCM). PCM merupakan model penskoran dari model Rasch yang dikembangkan menjadi skor politomus (Retnawati, 2014). Pemilihan model tersebut dilakukan karena menurut De Ayala pada PCM tidak mensyaratkan langkah penyelesaian harus berurutan dan tidak harus mempunyai kesulitan yang sama (Isgiyanto, 2013). Treshold pada penyekoran PCM juga tidak selalu lebih besar. Oleh karena itu PCM cocok untuk butir dengan model penilaian berjenjang dengan indeks kesukaran dalam setiap proses tidak perlu berurutan. Pengkategorian hasil tes dilakukan dengan memberikan skor pada jawaban siswa. Kriteria penilaian untuk tes AKM numerasi dengan bentuk teks pilihan ganda, pilihan ganda kompleks, menjodohkan, isian singkat, dan uraian dengan tipe data politomus. Selain kolom jawaban soal, dalam penelitian ini juga dikembangkan alasan dalam menjawab soal yang juga memberikan pertimbangan terhadap skor akhir.

### **C. Hasil dan Pembahasan**

Hasil pada penelitian ini dijelaskan melalui tiga bagian yang mempermudah pemahaman bacaan yakni terdiri dari konstruk instrumen, hasil uji kualitas instrumen, dan hasil asesmen.

## 1. Hasil

### a. Konstruksi instrumen

Instrumen tes numerasi dalam penelitian ini dikembangkan secara digital untuk dapat digunakan secara online atau melalui basis komputer. Produk akhir yang dikembangkan dalam penelitian ini adalah seperangkat instrumen penilaian numerasi yang terdiri dari kisi-kisi instrumen, instrumen, dan rubrik penskoran. Domain yang disertakan dalam numerasi ini adalah domain bilangan, aljabar, geometri dan pengukuran, serta data dan ketidakpastian. Sub domain dan kompetensi pada domain disesuaikan dengan *framework* AKM oleh Pusmenjar yakni fase B. Hal tersebut dilakukan dengan pertimbangan bahwa siswa yang telah selesai fase B adalah siswa kelas 4 yang sudah tuntas dimana mereka yang sudah masuk di awal kelas 5.

Soal yang dibuat merupakan soal dua tingkat (*two-tier*) dengan konteks stimulus mengenai kehidupan sehari-hari (*daily live activities*) seorang tokoh karakter utama pada instrumen soal tersebut. Tes *two tier* dimaksudkan agar lebih praktis digunakan oleh siswa sebagai peserta tes dan lebih bermakna bagi guru dalam hal mengurangi tebakan jawaban oleh siswa sehingga memungkinkan pengelolaan dan evaluasi yang lebih kredibel dan memberikan informasi tentang alasan siswa (Xiao et al., 2018). Terdapat 17 pendekatan stimulus pada konstruksi soal yang dibuat yang terdiri atas tiga konteks yakni personal, sosial budaya, dan saintifik. Matriks instrumen yang mencakup penjabaran domain, konteks, level kognitif, dan bentuk soal serta jumlah butir pada instrumen yang dikembangkan dapat dilihat pada tabel 1.

Tabel 1. Matriks Instrumen

		Bilangan	Aljabar	Geometri dan Pengukuran	Data dan Ketidakpastian	Jumlah
<b>Konteks</b>	Personal	5	2	2	3	40
	Sosial Budaya	3	3	2	3	
	Saintifik	-	3	6	4	
	<b>Level Kognitif</b>					
<b>Level Kognitif</b>	Pemahaman	1	1	3	1	40
	Penerapan	2	1	4	7	
	Penalaran	7	8	3	2	
<b>Bentuk Soal</b>	PG	1	3	4	3	40
	PGA	4	2	1	2	
	PGB	1	2	3	2	

	Bilangan	Aljabar	Geometri dan Pengukuran	Data dan Ketidakpastian	Jumlah
Menjodohkan	3	1	1	-	
Isian Singkat	1	1	-	2	
Uraian	-	1	2	1	

Konteks stimulus pada instrumen tersebut merupakan kegiatan seputar aktivitas peserta didik sekolah dasar yang dengan dihidirkannya satu tokoh sentral yang terlibat dalam semua aktivitas stimulus. Konteks stimulus tersebut terdiri atas 5 konteks personal, 6 konteks sosial budaya, dan 6 konteks saintifik. Stimulus pada instrumen ini digunakan untuk menjelaskan permasalahan pada butir soal, dalam instrumen ini terdiri dari 17 stimulus. Masing-masing stimulus mengampu 2-4 soal dengan bentuk soal yang berbeda-beda. Hal tersebut disesuaikan dengan ketersediaan pertanyaan yang dapat dikaji pada permasalahan stimulus terkait. Memperkaya pembelajaran matematika dengan stimulus aktivitas kehidupan sehari-hari (*everyday life activity stimulus*) menjadi kunci penting dalam menjembatani konsep abstrak matematika dengan penerapannya dalam konteks nyata. Pendekatan ini membantu siswa memahami bahwa matematika bukan hanya kumpulan rumus dan angka, melainkan alat yang berguna untuk menyelesaikan masalah dan memahami dunia di sekitar mereka. Stimulus tersebut mengandung konteks pendekatan secara personal, sosial budaya, dan saintifik. Stimulus pada instrumen yang dikembangkan merupakan kegiatan sehari-hari yang dilakukan oleh tokoh utama dengan nama Kamala.



*Gambar 2. Tokoh Utama Konteks Stimulus  
(dibuat oleh Hapsari dan Khairani dengan Adobe Illustrator)*

Kegiatan harian tokoh Kamala tersebut menjadi stimulus soal yang disesuaikan dengan konteks yang ada pada *framework* AKM. Berikut ini adalah contoh kegiatan yang dilakukan oleh Kamala dalam konteks stimulus personal domain bilangan dan subdomain representasi bilangan. Stimulus tersebut menjelaskan permasalahan personal Kamala mengenai uang tabungannya yang akan digunakan untuk keperluan sekolah. Kamala dibantu oleh Kinan kakaknya untuk menuliskan keperluan sekolah Kamala, namun Kinan menuliskannya dalam bentuk representasi bilangan pecahan, persentase, dan bilangan desimal. Melalui konteks stimulus tersebut peserta tes diharapkan dapat menghitung alokasi kebutuhan sekolah dengan representasi bilangan dengan kompetensi menggunakan operasi hitung bilangan cacah dan mengetahui posisi pecahan pada garis bilangan. Peserta tes dapat mengaitkan kemampuan numerasinya dengan persoalan matematis yang sering dihadapi oleh anak dalam konteks personalnya. Melalui soal ini, siswa difasilitasi untuk mendapatkan pengalaman dalam menentukan proporsi dan prioritas keperluan sekolah sehingga lebih berhati-hati dalam menggunakannya.



Gambar 3. Contoh Stimulus dalam Konteks Personal  
(dibuat oleh Hapsari dan Khairani melalui Adobe Illustrator)

Pengembangan instrumen ini menggunakan pedoman yang sesuai dengan *framework* AKM dan dikembangkan dengan prosedur pengembangan modifikasi penelitian Oriondo dan Antonio (1988) dan Mardapi (2017). Prosedur pengembangan tersebut melalui tiga tahapan utama yaitu (1) perancangan instrumen, (2) pengujicobaan instrumen, dan (3) pengukuran. Tahapan

perancangan instrumen tes dilakukan dengan langkah-langkah seperti: penentuan tujuan tes, penentuan kompetensi, penentuan konstruk materi, penyusunan kisi-kisi tes, dan penyusunan item tes. Kemudian pada tahapan pengujian instrumen tes dilakukan dua kali dengan subjek yang berbeda. Responden pada penelitian ini diambil dengan metode random sampling yang meliputi semua kecamatan yang dalam Kabupaten Kulon Progo disebut sebagai Kapanewon dengan sebaran data pada tabel 2.

Tabel 2. Matriks Instrumen

Kapanewon (Kecamatan)	Jumlah Sekolah				Jumlah Siswa	
	Negeri		Swasta		Uji 1	Uji 2
	Uji 1	Uji 2	Uji 1	Uji 2		
Wates	3	10	2	1	24	65
Panjatan	1	3	1	0	9	9
Temon	0	4	0	0	0	20
Galur	8	8	6	1	73	52
Lendah	3	2	2	0	19	13
Pengasih	2	7	0	1	9	29
Kokap	0	1	0	0	0	1
Sentolo	2	4	4	0	35	13
Nanggulan	2	2	0	1	9	20
Samigaluh	4	3	0	0	6	8
Girimulyo	2	4	1	0	16	12
Jumlah	27	48	16	4	200	242

Pengujian instrumen yang pertama dilakukan untuk pengujian validitas isi, validitas konstruk, dan reliabilitas instrumen sehingga didapatkan instrumen yang valid untuk diujicobakan pada uji coba yang kedua. Analisis yang digunakan pada uji coba yang kedua adalah analisis PCM dengan uji prasyarat respon butir dengan bantuan program R dan SPSS yang juga dimaksudkan untuk mengukur capaian numerasi siswa sekolah dasar.

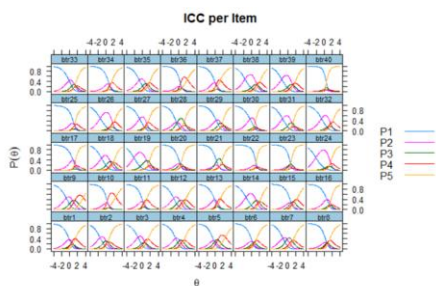
#### b. Hasil uji kualitas instrumen

Hasil validitas isi dari hasil penilaian yang diberikan oleh para ahli terhadap instrumen asesmen numerasi telah memenuhi kriteria valid berdasarkan hasil analisis validitas Aiken dengan nilai pada semua butirnya yakni lebih dari 0,7. Analisis validitas konstruk pada penelitian ini menggunakan analisis EFA (*Exploratory Factor Analysis*) yang dilakukan dengan

menggunakan bantuan program SPSS (*Statistical Package for Social Science*). Analisis *KMO Bartlett's* menunjukkan nilai sebesar 0,788 yang lebih dari 0,7, sehingga dapat dilakukan analisis faktor yang menghasilkan nilai *MSA (Measure of Sampling Adequacy)* dengan satu butir yang kurang dari 0,5 sehingga butir tersebut harus diperbaiki.

Secara keseluruhan hasil total *variance explained* memiliki persentase kumulatif sebesar 62,235% yang berarti faktor tersebut mampu mengukur kemampuan numerasi siswa sebesar 62,235%, dan sisanya diukur oleh faktor lain. Berdasarkan analisis *rotated component matrix* yang dilakukan terdapat 3 butir yang kurang valid karena memiliki *loading factor* yang di bawah 0,4. Butir tersebut adalah butir dengan nomor 7, 9, dan 11. Selain butir tersebut butir yang lain telah memenuhi kriteria valid sehingga dapat digunakan. Butir yang tidak valid akan dikeluarkan dari instrumen tes atau diperbaiki dari segi konstruksi materi yang mendasarinya. Hasil analisis reliabilitas menunjukkan instrumen numerasi yang digunakan dapat dikatakan reliabel karena nilai *Cronbach's Alpha* lebih dari 0,60 yaitu 0,872.

Selanjutnya dilakukan uji parameter butir yakni tingkat kesulitan soal atau (b). Hasil analisis PCM pada instrumen tes ini terdapat 4 tingkat kesulitan karena bentuk penskoran instrumen menggunakan skala 0, 1, 2, 3, dan 4. Tingkat kesulitan pertama (b1) berada pada kisaran -2,12 sampai +2,15. Tingkat kesulitan kedua (b2) berada pada kisaran -0,38 sampai dengan +2,39. Tingkat kesulitan ketiga (b3) berada pada kisaran -1,45. Sedangkan tingkat kesulitan keempat (b4) berada pada tingkat kisaran -1,09 sampai +3,39. Hasil analisis parameter butir menunjukkan grafik ICC (*Item Characteristic Curves*) sebagai berikut.



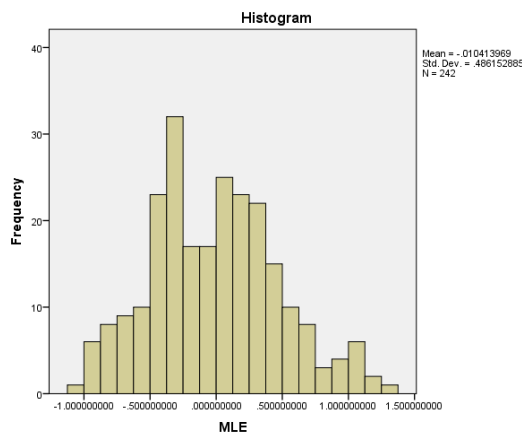
Gambar 4. Item characteristic curves

Setelah melalui proses analisis berikut ini disusun tabel kumulatif kategorisasi *threshold* pada instrumen numerasi yang dikembangkan.

Tabel 3. Kategorisasi Treshold

Kategori	b1	b2	b3	b4	Jumlah
Sangat Sukar	1	6	4	8	19
Sukar	5	16	18	21	60
Sedang	22	18	17	10	67
Mudah	10	0	1	1	12
Sangat Mudah	2	0	0	0	2

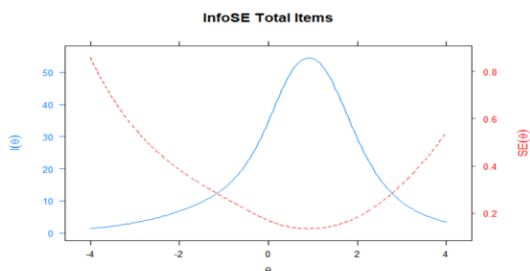
Tabel tersebut menunjukkan bahwa besaran *threshold* karakteristik butir menunjukkan bahwa instrumen yang dikembangkan memiliki kategori sedang dengan jumlah terbesar pertama dan sukar pada jumlah terbesar selanjutnya. Parameter kemampuan peserta tes yang digunakan pada analisis ini adalah metode MLE (*Maximum Likelihood Estimation*) dengan grafik sebagai berikut.



Gambar 5. Grafik MLE

Berdasarkan histogram tersebut diketahui bahwa paling banyak responden berada pada kemampuan -0,2 dan -0,3 dengan grafik cenderung ke kiri.

Selanjutnya dilakukan analisis fungsi informasi yang memberikan keterangan seberapa besarnya kontribusi butir soal dalam menjelaskan kemampuan peserta tes yang ditunjukkan dengan gambar sebagai berikut.



Gambar 7. Grafik TIF dan Standar Error

Berdasarkan analisis yang dilakukan dengan menggunakan parameter butir menunjukkan bahwa tes yang dikembangkan memberikan informasi maksimal sebesar 58 dengan standar error 0,7 jika diberikan pada peserta tes yang memiliki tingkat kemampuan sedang yaitu 1. Oleh karena itu, instrumen tes ini termasuk instrumen yang baik atau andal untuk mengukur kemampuan numerasi siswa. Batas bawah interval terdapat pada  $\theta = -1,2$  (responden dengan kemampuan rendah) dan batas atas interval terdapat pada  $\theta = 2,8$  (responden dengan kemampuan tinggi). Hasil tersebut menunjukkan bahwa tes baik diberikan pada kemampuan  $-1,2$  sampai dengan  $+2,8$ . Hal tersebut karena kesalahan pengukuran yang dilakukan pada interval tersebut tidak semakin besar. Instrumen dikatakan reliabel digunakan pada peserta tes dengan tingkat kemampuan rendah ( $-1,7$ ) sampai dengan kategori tinggi ( $+2,8$ ).

### c. Hasil asesmen

Hasil pengerjaan tes yang diinterpretasi merupakan hasil tes pada uji coba yang ke-2 dengan mempertimbangkan butir yang tidak fit pada hasil analisis PCM dan konstruksi kisi-kisi instrumen. Oleh karena itu hasil interpretasi tes yang dilakukan melibatkan 242 responden dengan 34 butir instrumen. Ketercapaian numerasi disajikan dengan mengubah kemampuan siswa dari skala output R program yang ke dalam skala 0 hingga 100. Hal tersebut dilakukan untuk mempermudah penyajian dan penafsiran tingkat kemampuan yang lazim dikenal di dunia pendidikan.

Berikut ini adalah kategorisasi kemampuan numerasi yang dikelompokkan ke dalam 5 kategori dengan penghitungan berdasarkan simpangan bakunya ( $SB$ ) dan rata-ratanya ( $x$ ).

Tabel 4. Kategorisasi kemampuan Numerasi

Kemampuan Numerasi	Kategori	Jumlah	Persentase
$X \geq \bar{x} + 1,5 S$	Mahir	25	10,33%
$\bar{x} \leq X < \bar{x} + 1,5 S$	Cakap	70	28,93%
$\bar{x} - 1,5 SB \leq X < \bar{x}$	Dasar	146	60,33%
$X < \bar{x} - 1,5 S$	Perlu Intervensi Khusus	1	0,41%

## 2. Pembahasan

Instrumen numerasi yang dikembangkan diperkaya menggunakan konteks stimulus yang berbeda dengan yang selama ini ada. Konteks tersebut merupakan kegiatan seputar aktivitas peserta didik sekolah dasar yang dengan dihadapkannya satu tokoh sentral yang terlibat dalam semua aktivitas stimulus. Konteks stimulus tersebut terdiri atas 5 konteks personal, 6 konteks sosial budaya, dan 6 konteks saintifik. Stimulus pada instrumen ini digunakan untuk menjelaskan permasalahan pada butir soal, dalam instrumen ini terdiri dari 17 stimulus. Masing-masing stimulus digunakan dalam 2-4 soal dengan bentuk soal yang berbeda-beda. Hal tersebut disesuaikan dengan ketersediaan pertanyaan yang dapat dikaji pada permasalahan stimulus terkait.

Stimulus yang digunakan mengandung konteks pendekatan secara personal, sosial budaya, dan saintifik. Stimulus pada instrumen yang dikembangkan merupakan kegiatan sehari-hari atau *everyday life activity* yang dilakukan oleh tokoh utama dengan nama Kamala. Stimulus dengan konteks sehari-hari berbantuan gambar dapat menambah kemampuan siswa dalam penyelesaian masalah (Novotná et al., 2012). *Everyday life activity* juga dapat membantu siswa lebih mudah mengembangkan kemampuan penyelesaian masalah (Sari et al., 2023). Siswa dalam memecahkan masalah tidak hanya bergantung pada hafalan, tetapi juga menganalisis informasi (Mestre et al., 2011) dan berpikir kritis untuk menelaah masalah dalam *everyday life activity* (Syafri et al., 2020). Stimulus pertanyaan dapat berfungsi dengan baik. Konteks yang dipilih mudah dipahami oleh siswa karena berhubungan langsung dengan kegiatan kesehariannya. Hal tersebut dibuktikan dengan tidak adanya komplain mengenai gambar, simbol, ataupun grafik dalam soal. Sebelum uji coba yang kedua terdapat beberapa stimulus yang

diperbaiki sesuai dengan kebutuhan permasalahan dalam soal. Keterbacaan stimulus juga dapat dibaca dengan baik.

Instrumen tes yang baik dan layak dapat digunakan sesuai dengan peruntukannya. Melalui *every day life activity* stimulus siswa yang didukung dengan bentuk two tier siswa dapat lebih memahami persoalan dengan bentuk soal yang berbeda-beda dan menyebutkan alasan mengapa siswa memilih jawaban yang demikian. Disamping itu, penggunaan alasan saat menjawab *two tier* tes adalah cara yang sensitif dan efektif untuk menambah nilai kelengkapan item pertanyaan pilihan jawaban mereka (Abidin & Retnawati, 2019). Instrumen numerasi yang dikembangkan telah layak digunakan untuk menstimulasi penalaran siswa untuk menyelesaikan permasalahan sehari-hari secara matematis. Suatu instrumen yang efektif dapat dilihat pada pengukuran mencapai aspek indikator yang telah ditetapkan (Maulidiansyah et al., 2021). Hasil analisis yang menunjukkan bahwa instrumen tes yang dikembangkan mampu mengidentifikasi tingkat kemampuan numerasi siswa.

Instrumen numerasi untuk siswa sekolah dasar ini dapat digunakan sebagai sarana diagnosis kelemahan kemampuan numerasi siswa dalam menyelesaikan tugas belajarnya. Instrumen diagnostik dapat digunakan untuk melakukan intervensi yang efektif kepada siswa secara individual ataupun klasikal, dalam upaya mengevaluasi proses pembelajaran (Duskri et al., 2014). Hal tersebut sesuai dengan konteks pada saat instrumen ini diujicobakan. Instrumen numerasi ini telah melewati proses pengembangan instrumen dengan berbagai perbaikan sehingga layak digunakan. Sejalan pendapat Faizah et al. (2019) yang menjelaskan bahwa proses pengembangan instrumen yang mengikuti prosedur pengembangan yang sistematis akan menghasilkan produk yang layak digunakan. Pengembangan instrumen ini juga berguna untuk melatih daya pikir matematis peserta didik yang dapat digunakan dalam kehidupan sehari-harinya. Sehingga peserta didik akan lebih memahami literasi numerasi yang aplikatif.

Kualitas instrumen numerasi ini dilihat dari empat aspek yaitu tingkat validitas isi, validitas konstruk, reliabilitas instrumen, dan kualitas butir. Selanjutnya instrumen AKM numerasi juga diuji secara empiris untuk

menganalisis validitas konstruknya. Validitas konstruk merupakan validitas yang menjelaskan sejauh mana item tes dapat mengukur apa yang benar-benar ingin diukur berdasarkan konsep tertentu atau definisi konsep yang ditentukan (Sustekova et al., 2019). Uji kualitas butir dilakukan dengan analisis PCM (Partial Credit Model). Model PCM adalah model penskoran politomus yang merupakan perluasan dari model rasch (Retnawati, 2016). PCM berfokus pada kategori yang berdekatan ketika PCM mampu memberikan informasi mengenai tingkat kesukaran siswa yang diperoleh melalui kesalahan mereka pada setiap tingkat kesulitan yang berbeda (Shantika & Istiyono, 2019). Asumsi ini membuktikan bahwa setiap butir tes hanya mengukur satu kemampuan. Unidimensi pada instrumen numerasi ini dapat dilihat dari nilai eigen yang membentuk faktor dalam uji analisis faktor sebelumnya (Kartowagiran et al., 2019). Penggunaan PCM pada instrumen numerasi yang dikembangkan diharapkan mampu memberikan informasi yang lebih baik untuk mengetahui tingkat kemampuan siswa. Berdasarkan hasil analisis yang dilakukan, instrumen memberikan informasi maksimal sebesar 58 dengan standard error 0,7 jika diberikan pada peserta tes yang memiliki kemampuan sedang yaitu 1. Nilai fungsi informasi adalah cara untuk menjelaskan besarnya kontribusi butir soal dalam mengungkap kemampuan peserta tes (Safaruddin et al., 2012). Fungsi informasi digunakan untuk menentukan item yang cocok untuk model, sehingga item dapat dipilih dengan lebih mudah. Fungsi informasi pengujian adalah jumlah dari semua fungsi informasi proyek yang menyusun pengujian.

Batas bawah interval terdapat pada teta -1,2 (responden dengan kemampuan rendah) dan batas atas interval terdapat pada teta 2,8 (responden dengan kemampuan tinggi) memperkirakan ambang kesulitan antara kategori tanggapan yang ada. Hasil uji coba instrumen numerasi yang dianalisis merupakan tahap uji coba instrumen yang ke-2 dengan 242 responden. Tingkat kemampuan siswa dikategorikan melalui empat kategori yang disarankan oleh Kemendikbud (2020) yakni mahir, cakap, dasar dan perlu intervensi khusus. Berdasarkan hasil analisis yang dilakukan diketahui bahwa sebaran kemampuan numerasi responden terdiri atas 10,3% responden

dengan kemampuan mahir, 28,3% responden dengan kemampuan cakap, 60,3% responden dengan kemampuan dasar, dan 0,4% responden yang memerlukan intervensi khusus.

Melalui analisis masing-masing domain numerasi didominasi oleh kemampuan siswa yang berada pada kategori dasar. Siswa masih belum mampu menerapkan konsep matematis terhadap suatu permasalahan yang dihadapi walaupun telah memiliki kemampuan dasar matematis. Hal yang dapat dilakukan oleh lingkungannya baik guru ataupun orang tua adalah memberikan contoh-contoh permasalahan sehari-hari yang relevan terhadap kehidupan siswa yang dapat diselesaikan secara matematis. Peningkatan kemampuan matematis siswa tersebut dapat ditingkatkan melalui beberapa strategi diantaranya adalah penggunaan bahan ajar realistik (Ulandari et al., 2019), penerapan *project based learning* dan *guided discovery learning* (Supriadi et al., 2019), dan lain sebagainya yang disesuaikan dengan karakteristik siswa pada tiap fasenya.

Setelah melalui uji coba dan analisis signifikansi kenaikan level numerasi juga ditunjukkan pada hasil AKM di tahun berikutnya yang naik 16,01% dari capaian tahun 2021 yakni sebanyak 30,66% menjadi 46,67% pada tahun 2022. Hal tersebut berarti kontekstualisasi persoalan matematis dengan *every day live activity* dapat membantu siswa untuk memahami konsep numerasi dan menerapkannya. Siswa juga dapat dibiasakan untuk mengurangi tebakan dalam menjawab soal melalui soal dengan bentuk *two tier. Framework* AKM sudah memberikan pengalaman bagi siswa dalam mengerjakan berbagai bentuk soal sehingga dimungkinkan bagi siswa untuk dapat mencari solusi dengan berbagai cara.

#### **D. Penutup**

Hasil penelitian menunjukkan bahwa model *two-tier test* berbasis stimulus karakter dengan konteks *every day life activity* secara signifikan dapat meningkatkan kemampuan numerasi siswa sekolah dasar. Hal tersebut terbukti dari perbandingan ketercapaian numerasi dari tahun sebelum penggunaan tes dengan tahun setelah penggunaan tes yang mengalami

peningkatan. Instrumen numerasi telah memiliki kualitas tes yang baik. Instrumen yang dikembangkan ini telah memenuhi kriteria valid, reliabel, dan memiliki kualitas butir soal yang baik. Temuan penelitian ini menunjukkan bahwa model *two tier test* merupakan pendekatan yang menjanjikan untuk meningkatkan kemampuan pada siswa sekolah dasar.

Melalui *two-tier test* siswa difasilitasi untuk dapat memberikan alasan spesifik dan cara penyelesaian untuk memberikan jawaban dengan berbagai bentuk soal. Konteks *every day life activity* juga memberikan gambaran nyata keterhubungan konsep matematis dengan penerapannya sehari-hari. *Framework* AKM yang memberikan berbagai pilihan bentuk soal juga mendukung eksplorasi pengalaman penyelesaian permasalahan yang beragam. Pembiasaan mengerjakan soal *two tier* akan mengurangi kebiasaan menebak oleh siswa sehingga dan secara terbuka juga menyediakan kesempatan lebih banyak bagi siswa untuk menunjukkan pemahaman dan kemampuan dalam menerima umpan balik yang dapat mengarah pada peningkatan kualitas pembelajaran.

Hasil penelitian ini dapat lebih lanjut dikembangkan pada mata pelajaran lain atau pada konteks yang lebih spesifik sesuai dengan tujuan dan kompetensi pembelajaran. Implikasi dari penelitian ini kiranya dapat menjadi pertimbangan bagi pihak-pihak yang memiliki kewenangan dalam hal penyusunan soal di tingkat sekolah ataupun lingkup yang lebih luas.

### **Ucapan Terima Kasih**

Ucapan terima kasih peneliti sampaikan kepada Kepala Dinas Pendidikan Pemuda dan Olahraga Kabupaten Kulon Progo, forum kerja Pengawas Sekolah Dasar, guru dan siswa sebagai responden yang telah berkenan memberikan bantuan dan berpartisipasi dalam pelaksanaan penelitian ini.

### **Daftar Referensi**

Abidin, M., & Retnawati, H. (2019). A Diagnosis of Difficulties in Answering Questions of Circle Material on Junior High School Students. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 23(2), 144-155. <https://doi.org/10.21831/pep.v23i2.16454>.

- Achmetli, K., Schukajlow, S., & Rakoczy, K. (2019). Multiple Solutions for Real-World Problems, Experience of Competence and Students' Procedural and Conceptual Knowledge. *International Journal of Science and Mathematics Education*, 17(8), 1605–1625. <https://doi.org/10.1007/s10763-018-9936-5>
- Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>.
- Allen, M. ., & Yen, W. (1979). *Introduction to Measurement Theory*. Wadsworth.Inc.
- Asrizal, A., Amran, A., Ananda, A., & Festiyed, F. (2018). Effectiveness of Adaptive Contextual Learning Model of Integrated Science by Integrating Digital Age Literacy on Grade VIII Students. *IOP Conference Series: Materials Science and Engineering*, 335(1). <https://doi.org/10.1088/1757-899X/335/1/012067>.
- Azhari, B., Johar, R., Ramadhani, E., Mailizar, M., & Safrina, K. (2024). Mathematics Learning Model for Children with Dyscalculia through Special Intervention. *Jurnal Ilmiah Peuradeun*, 12(3), 1155–1184. <https://doi.org/10.26811/peuradeun.v12i3.1528>.
- Aziza, M. (2019). Kemampuan Pemecahan Masalah Siswa dalam Menyelesaikan Soal Tertutup dan Terbuka pada Pokok Bahasan Lingkaran. *Pythagoras: Jurnal Pendidikan Matematika*, 14(2), 126–138. <https://doi.org/10.21831/pg.v14i2.26563>.
- Bennison, A., Goos, M., & Geiger, V. (2020). Utilising a Research-Informed Instructional Design Approach to Develop an Online Resource to Support Teacher Professional Learning on Embedding Numeracy Across the Curriculum. *ZDM - Mathematics Education*, 52(5), 1017–1031. <https://doi.org/10.1007/s11858-020-01140-2>.
- Clarke, B. (2022). Research in mathematics education in Australasia 2016–2019. In *Research in Mathematics Education*, 24(1). <https://doi.org/10.1080/14794802.2021.1899971>.
- Cobo, C. (2013). Skills for Innovation: Envisioning an Education that Prepares for the Changing World. *Curriculum Journal*, 24(1), 67–85. <https://doi.org/10.1080/09585176.2012.744330>.
- Dole, S., & Geiger, V. (2020). *Numeracy Across the Curriculum: Research-Based Strategies for Enhancing Teaching and Learning*. Routledge.

- Duskri, M., Kumaidi, K., & Suryanto, S. (2014). Pengembangan Tes Diagnostik Kesulitan Belajar Matematika di SD. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 18(1), 44–56. <https://doi.org/10.21831/pep.v18i1.2123>.
- Faizah, U., Zuchdi, D., & Alsamiri, Y. (2019). An Authentic Assessment Model to Assess Kindergarten Students' Character. *Research and Evaluation in Education*, 5(2), 103–119. <https://doi.org/10.21831/reid.v5i2.24588>.
- Geiger, V., Forgasz, H., & Goos, M. (2015). A Critical Orientation to Numeracy Across the Curriculum. *ZDM - International Journal on Mathematics Education*, 47(4), 611–624. <https://doi.org/10.1007/s11858-014-0648-1>.
- Goos, M., Geiger, V., & Dole, S. (2011). Teachers' Personal Conceptions of Numeracy. *35th Conference of the International Group for the Psychology of Mathematics Education*, 2, 457–464.
- Halim, A., Mustafa, Nurulwati, Soewarno, & Nanda, N. (2018). Development of Two-Tier Diagnostic Test Based on E-Learning. *Journal of Physics: Conference Series*, 1120(1). <https://doi.org/10.1088/1742-6596/1120/1/012030>.
- Hamidy, A., & Jailani, J. (2019). Kemampuan Proses Matematis Siswa Kalimantan Timur dalam menyelesaikan soal matematika model PISA. *Jurnal Riset Pendidikan Matematika*, 6(2), 133–149. <https://doi.org/10.21831/jrpm.v6i2.26679>.
- Haryanto, H. (2013). Pengembangan Computerized Adaptive Testing (Cat) Dengan Algoritma Logika Fuzzy. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(1), 47–70. <https://doi.org/10.21831/pep.v15i1.1087>.
- Isgiyanto, A. (2013). Perbandingan Penyekoran Model Rasch dan Model Partial Credit pada Matematika. *Jurnal Kependidikan*, 43(1), 9–18. <https://doi.org/10.21831/jk.v43i1.1954>.
- Jupri, A., Drijvers, P., & Van den Heuvel-Panhuizen, M. (2014). Student Difficulties in Solving Equations from an Operational and a Structural Perspective. *International Electronic Journal of Mathematics Education*, 9(1–2), 39–55. <https://doi.org/10.29333/iejme/280>
- Kartowagiran, B., Mardapi, D., Purnama, D. N., & Kriswantoro, K. (2019). Parallel Tests Viewed from the Arrangement of Item Numbers and Alternative Answers. *Research and Evaluation in Education*, 5(2), 169–182. <https://doi.org/10.21831/reid.v5i2.23721>.
- Kemendikbud. (2020). AKM dan Implikasinya pada Pembelajaran. *Pusat Asesmen Dan Pembelajaran Badan Penelitian Dan Pengembangan Dan*

*Perbukuan Kementerian Pendidikan Dan Kebudayaan Pembelajaran Badan Penelitian Dan Pengembangan Dan Perbukuan Kementerian Pendidikan Dan Kebudayaan, 1–37.*

- Khasanah, D. R. A. U., Pramudibyanto, H., & Widuroyekti, B. (2020). Pendidikan dalam Masa Pandemi Covid-19. *Jurnal Sinestesia*, 10(1), 41–48. <https://www.sinestesia.pustaka.my.id/journal/article/view/44>.
- Khoiriah, K. (2022). Assessment for Learning Berorientasi Higher Order Thinking Skills untuk Menstimulus Kecakapan Literasi Numerasi. *Jurnal Didaktika Pendidikan Dasar*, 6(1), 127–144. <https://doi.org/10.26811/didaktika.v6i1.740>.
- Kilpatrick, J., Swafford, J., & Findell, B. (2001). Helping Children Learn Mathematics. In *Center for Education Division of Behavioral and Social Sciences and Education National Research Council* (Vol. 17, Issue 12). National Academy Press.
- Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S. B., Dale, P. S., & Plomin, R. (2013). Literacy and Numeracy Are More Heritable Than Intelligence in Primary School. *Psychological Science*, 24(10), 2048–2056. <https://doi.org/10.1177/0956797613486982>.
- Li, F. Y., Hwang, G. J., Chen, P. Y., & Lin, Y. J. (2021). Effects of a Concept Mapping-Based Two-Tier Test Strategy on Students' Digital Game-Based Learning Performances and Behavioral Patterns. *Computers and Education*, 173(June), 104293. <https://doi.org/10.1016/j.compedu.2021.104293>.
- Mardapi, D. (2007). *Teknik Penyusunan Instrumen Tes dan NonTes*. Mitra Cendekia.
- Maulidiansyah, D., Meutia, I., & Istiyono, E. (2021). Computer-Based Two-Tier Diagnostic Test to Identify Critical Thinking Skills in Optical Instrument. *Proceedings of the 6th International Seminar on Science Education (ISSE 2020)*, 541(Isse 2020), 413–418. <https://doi.org/10.2991/assehr.k.210326.059>.
- Md-Ali, R., Karim, H. B. B. A., & Yusof, F. M. (2016). Experienced Primary School Teachers' Thoughts on Effective Teachers of Literacy and Numeracy. *Malaysian Journal of Learning and Instruction*, 13(1), 43–62. <https://doi.org/10.32890/mjli2016.13.1.3>.
- Mestre, J. P., Docktor, J. L., Strand, N. E., & Ross, B. H. (2011). Conceptual Problem Solving in Physics. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 55). Elsevier Inc.

<https://doi.org/10.1016/B978-0-12-387691-1.00009-0>.

- Munayati, Z., zulkardi, & Santoso, B. (2015). Kajian Soal Buku Teks Matematika Kelas X Kurikulum 2013 Menggunakan Framework PISA. *Jurnal Pendidikan Matematika Sriwijaya*, 9(2), 188–206. <https://unej.id/1cWvAd2>.
- Muzaffar, A., Irfan, A., & Tabrani ZA. (2020). Kemampuan Pedagogical Content Knowledge Alumni Pendidikan Bahasa Arab Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry Banda Aceh. *Jurnal Ilmiah Didaktika: Media Ilmiah Pendidikan dan Pengajaran*, 21(1), 41-60. doi: <http://dx.doi.org/10.22373/jid.v21i1.7129>.
- Novotná, J., Eisenmann, P., Příbyl, J., Ondrušová, J., & Břehovský, J. (2012). Problem Solving in School Mathematics Based on Heuristic Strategies. *Journal on Efficiency and Responsibility in Education and Science*, 7(1), 1–6. <https://doi.org/10.7160/eriesj.2013.070101.Introduction>.
- Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming Racially Literate About Data and Data-Literate About Race: Data Visualizations in the Classroom as a Site of Racial-Ideological Micro-Contestations. *Cognition and Instruction*, 34(4), 361–388. <https://doi.org/10.1080/07370008.2016.1210418>.
- Phillips, J. J., & Phillips, P. P. (2016). Handbook of Training Evaluation and Measurement Methods. In *Handbook of Training Evaluation and Measurement Methods*. <https://doi.org/10.4324/9780080572659>.
- Putri, R., Johar, R., & Munzir, S. (2022). Teachers' Perception About Islamic Values Integration Into Mathematics Learning Through Comics. *Jurnal Ilmiah Peuradeun*, 10(1), 105–116. <https://doi.org/10.26811/peuradeun.v10i1.606>
- Rennie, L., & Parker, L.. (1996). Placing Physics Problems in Real-Life Context: Students' Reactions and Performance. *Australian Science Teachers Journal*, 42(1), 55–59. <https://search.informit.org/doi/abs/10.3316/aeipt.71579>.
- Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya: untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Nuha Medika.
- Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian*. Parama.
- Safaruddin, Anisa, & AF, M. S. (2012). Partial Credit Model ( PCM ) dalam

- Penskoran Politomi pada Teori Respon Butir. *Jurnal Matematika, Statistika, & Komputasi*, 9(1), 39–48. <https://doi.org/10.20956/jmsk.v9i1.3397>
- Sari, I. P. M., Jatmiko, B., & Suprpto, N. (2023). Students' Physics Problem-Solving Skills in Daily Life Context: Between Confession and Fact. *Jurnal Penelitian Pendidikan IPA*, 9(1), 231–241. <https://doi.org/10.29303/jppipa.v9i1.2561>.
- Shahbari, J. A., Daher, W., Baya'a, N., & Jaber, O. (2020). Prospective Teachers' Development of Meta-Cognitive Functions in Solving Mathematical-Based Programming Problems With Scratch. *Symmetry*, 12(9). <https://doi.org/10.3390/SYM12091569>.
- Shantika, E. G., & Istiyono, E. (2019). A diagnosis of students' errors in answering the mathematics test in senior high school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 23(2), 129–143. <https://doi.org/10.21831/pep.v23i2.16370>.
- Supriadi, N., Syazali, M., Lestari, B. D., Dewi, E. S., Utami, L. F., Mardani, L. A., & Putra, F. G. (2019). The Utilization of Project Based Learning and Guided Discovery Learning: Effective Methods to Improve Students' Mathematics Ability. *Al-Ta Lim Journal*, 25(3), 261–271. <https://doi.org/10.15548/jt.v25i3.487>.
- Sustekova, E., Kubiátko, M., & Usak, M. (2019). Validation of critical thinking test on Slovak conditions. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(12). <https://doi.org/10.29333/ejmste/112295>.
- Syafril, S., Aini, N. R., Netriwati, Pahrudin, A., Yaumas, N. E., & Engkizar. (2020). Spirit of Mathematics Critical Thinking Skills (CTS). *Journal of Physics: Conference Series*, 1467(1). <https://doi.org/10.1088/1742-6596/1467/1/012069>.
- Ulandari, L., Amry, Z., & Saragih, S. (2019). Development of Learning Materials Based on Realistic Mathematics Education Approach to Improve Students' Mathematical Problem Solving Ability and Self-Efficacy. *International Electronic Journal of Mathematics Education*, 14(2), 375–383. <https://doi.org/10.29333/iejme/5721>.
- Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word Problems in Mathematics Education: A Survey. *ZDM - Mathematics Education*, 52(1), 1–16. <https://doi.org/10.1007/s11858-020-01130-4>.

- Walidin, W., Idris, S., & Tabrani ZA. (2015). *Metodologi Penelitian Kualitatif & Grounded Theory*. FTK Ar-Raniry Press.
- 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), 20104. <https://doi.org/10.1103/PhysRevPhysEducRes.14.020104>.
- Yang, T. C., Chen, S. Y., & Hwang, G. J. (2015). The Influences of a Two-Tier Test Strategy on Student Learning: A Lag Sequential Analysis Approach. *Computers and Education*, 82, 366–377. <https://doi.org/10.1016/j.compedu.2014.11.021>.
- Zahroh, H., Hafidah, Dhofir, & Zayyadi, M. (2020). Gerakan Literasi Matematika dalam Peningkatan Kemampuan Pemecahan Masalah Matematis Siswa. *Sustainability (Switzerland)*, 4(1), 1–9. <https://doi.org/10.33387/dpi.v9i2.2293>.
- Zanabazar, A., Deleg, A., Ravdan, M., & Tsogt-erdene, E. (2023). The Relationship between Mathematics Anxiety and Mathematical Performance among Undergraduate Students. *Jurnal Ilmiah Peuradeun*, 11(1), 309–322. <https://doi.org/10.26811/peuradeun.v11i1.780>.
- Ziaurrahman, Z., Tabrani ZA & Andriansyah, A. (2024). Pengembangan E-Book Interaktif untuk Menunjang Pembelajaran Diferensiasi pada Mata Pelajaran Pendidikan Agama Islam di Sekolah Dasar. *Jurnal Didaktika Pendidikan Dasar*, 8(1), 165–184. <https://doi.org/10.26811/didaktika.v8i1.1333>.

