

## **PENGEMBANGAN INSTRUMEN TES NUMERASI BERBASIS TIMSS BAGI SISWA KELAS IV SEKOLAH DASAR**

Nur Rohmah<sup>1</sup>, Wasino<sup>2</sup>, Agus Yuwono<sup>3</sup>

<sup>123</sup>Universitas Negeri Semarang

<sup>1</sup>[nurrohmah17@students.unnes.ac.id](mailto:nurrohmah17@students.unnes.ac.id), <sup>2</sup>[wasino@mail.unnes.ac.id](mailto:wasino@mail.unnes.ac.id),

<sup>2</sup>[agusyuwono@mail.unnes.ac.id](mailto:agusyuwono@mail.unnes.ac.id)

### **ABSTRACT**

*This study aims to develop a TIMSS-based numeracy test instrument for fourth grade elementary school students through the stages of development research. This study also aims to prove the validity and test the reliability of the instrument, analyze the characteristics of the test instrument based on item responses, analyze the practicality and determine the numeracy profile of test takers. This research is a type of Research and Development (R&D with reference to the Four-D (4D) model, which consists of four stages. In the process of developing test instruments, student involvement was carried out in stages, namely, 19 students in the small-scale test trial and 90 students in the large-scale test trial stage. Data collection was carried out through several methods, such as interviews, distributing questionnaires, and giving tests. Data analysis was carried out using qualitative and quantitative approaches, where qualitative data was obtained from the results of the needs analysis, while quantitative data was obtained through expert assessments, the results of small-scale trials, and large-scale trials. The results showed that the TIMSS-based numeracy test instrument had content validity >0.80 on each item and instrument reliability level of 0.72. The value of agreement between experts based on the Intraclass Correlation Coefficient (ICC) test also showed a figure of 0.72. Based on item analysis using the Rasch model, test items are distributed in four levels of difficulty, namely very difficult questions of 13.3%, difficult 40%, easy 30% and very easy 16.7%. Overall, this instrument meets the item eligibility criteria and is considered suitable for use in measuring the numeracy skills of fourth grade elementary school students.*

*Keywords: Test Instrument Development, Numeracy, TIMSS, Elementary School*

### **ABSTRAK**

Penelitian ini bertujuan untuk mengembangkan instrumen tes numerasi berbasis TIMSS bagi siswa kelas IV sekolah dasar melalui tahapan penelitian pengembangan. Penelitian ini juga bertujuan untuk membuktikan validitas dan menguji reliabilitas instrumen, melakukan analisis terhadap karakteristik instrumen tes berdasarkan respons butir soal, melihat kepraktisan instrumen dan mengetahui profil numerasi peserta tes. Penelitian ini merupakan jenis penelitian *Research and Development* (R&D) dengan mengacu pada model Four-D (4D), yang terdiri dari empat tahapan. Dalam proses pengembangan instrumen tes, keterlibatan siswa dilakukan secara bertahap, yaitu, 19 siswa dalam uji coba tes skala kecil dan 90

siswa pada tahap uji coba tes skala besar. Pengumpulan data dilakukan melalui beberapa metode, seperti wawancara, penyebaran kuesioner, dan pemberian tes. Analisis data dilakukan dengan pendekatan kualitatif dan kuantitatif, di mana data kualitatif diperoleh dari hasil analisis kebutuhan, sedangkan data kuantitatif diperoleh melalui penilaian ahli, hasil uji coba skala kecil, serta uji coba skala besar. Hasil penelitian menunjukkan bahwa instrumen tes numerasi berbasis TIMSS memiliki validitas isi  $>0.80$  pada setiap butir dan tingkat reliabilitas instrumen sebesar  $0,72$ . Nilai kesepakatan antar ahli berdasarkan uji *Intraclass Correlation Coefficient* (ICC) juga menunjukkan angka sebesar  $0,72$ . Berdasarkan analisis butir menggunakan model Rasch, butir-butir tes terdistribusi dalam empat tingkat kesukaran, yaitu soal sangat sulit sebesar  $13,3\%$ , sulit  $40\%$ , mudah  $30\%$  dan sangat mudah  $16,7\%$ . Secara keseluruhan, instrumen ini memenuhi kriteria kelayakan butir dan dinilai sesuai untuk digunakan dalam mengukur kemampuan numerasi siswa kelas IV SD.

**Kata Kunci:** Pengembangan Instrumen Tes, Numerasi, TIMSS, Sekolah Dasar

### **A. Pendahuluan**

Kemampuan numerasi merupakan kemampuan dasar yang sangat penting bagi setiap individu. Kemampuan ini tidak hanya dibutuhkan dalam konteks pendidikan, tetapi juga dalam kehidupan sehari-hari (Tenny et al., 2021a). Terdapat beberapa literatur yang menunjukkan keterkaitan kecakapan numerasi di berbagai aspek kehidupan, mulai dari kesempatan kerja, besaran gaji (Dougherty, 2003), dan status sosial ekonomi (Gerardi et al., 2013), hingga hasil kesehatan fisik dan mental jangka panjang (Gross et al., 2009). Kecakapan numerasi juga diakui sebagai hal yang penting bagi

keberhasilan ekonomi suatu negara (Foley et al., 2017).

Kemampuan numerasi berhubungan erat dengan kemampuan matematika secara keseluruhan. Tiga materi dasar matematika yang menjadi aspek penting dari kemampuan numerasi adalah bilangan, geometri dan penyajian data. Pemahaman yang mendalam terhadap materi matematika memungkinkan siswa untuk tidak hanya menghafal rumus, tetapi juga memahami bagaimana dan kapan menerapkannya dalam situasi nyata. Hal ini sejalan dengan pendapat (Maulidina, 2019) yang menyatakan bahwa pemahaman matematika yang baik dapat mengarahkan siswa untuk memiliki

kemampuan numerasi yang baik. Kemampuan numerasi mencakup keterampilan berpikir menggunakan konsep, fakta, dan prosedur matematika untuk menyelesaikan masalah sehari-hari.

Siswa yang memahami dasar-dasar numerasi, seperti membilang, memahami nilai tempat, dan operasi hitung dasar, akan lebih mudah dalam mempelajari konsep matematika yang lebih tinggi, seperti aljabar dan geometri (Adinda et al., 2022). Oleh karena itu, perlu adanya upaya bersama antara guru, orang tua, dan lembaga pendidikan untuk meningkatkan kemampuan numerasi siswa sejak dini.

Indonesia saat ini sedang mengalami masalah serius akibat rendahnya kompetensi numerasi pada siswa. Hal ini terlihat dari banyaknya siswa, terutama di tingkat sekolah dasar yang masih mengalami kesulitan dalam memahami konsep matematika dan hanya mampu menyelesaikan soal-soal matematika sederhana (Prastyo, 2020). Pernyataan tersebut sejalan dengan hasil capaian kemampuan numerasi tingkat SD/MI/Sederajat di Indonesia pada Asesmen Nasional tahun 2022

menunjukkan sebesar 53,33% siswa memiliki kompetensi dibawah minimum (Kemendikbud, 2023). Kemudian pada tahun 2023 capaian dibawah minimum meningkat menjadi 57,46% (Kemendikbud, 2024). Artinya, lebih dari setengah siswa SD/ MI/ sederajat di Indonesia yang memiliki kompetensi numerasi rendah.

Peningkatan kemampuan numerasi siswa dapat dilakukan dengan penilaian yang baik dan valid. Penilaian merupakan proses pengumpulan dan pengolahan data untuk mengukur pencapaian hasil belajar siswa. Tujuan penilaian menurut Permendikbud Nomer 23 Tahun 2016 adalah untuk: (1) memantau dan mengevaluasi proses, perkembangan, serta peningkatan hasil belajar secara terus-menerus, (2) mengukur pencapaian standar kompetensi lulusan di semua mata pelajaran, dan (3) menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu.

Salah satu bentuk penilaian adalah dengan tes. Penilaian dengan tes sangat penting untuk mengidentifikasi kekuatan dan kelemahan siswa dalam bidang

numerasi, sehingga dapat dilakukan tindakan perbaikan yang tepat (Tenny et al., 2021b). Tes yang tepat dapat membantu guru mengetahui kemampuan siswa secara individual. Hasil tes memberikan informasi yang diperlukan untuk merancang intervensi pendidikan yang sesuai.

Saat ini terdapat penilaian internasional yang mengevaluasi kemampuan dan keahlian siswa dalam bidang matematika dan sains, yaitu *Trend in International Mathematics and Science Study* (TIMSS). TIMSS adalah studi internasional yang bertujuan untuk mengukur pencapaian siswa di Sekolah Dasar (SD) kelas empat dan Sekolah Menengah Pertama (SMP) kelas delapan dalam bidang matematika dan sains setiap empat tahun sekali. Hasil TIMSS tahun 2015 menunjukkan bahwa Indonesia berada di peringkat 44 dari 49 negara peserta, dengan skor rata-rata kemampuan matematika siswa hanya mencapai 397. Skor ini jauh di bawah standar TIMSS Scale Centerpoint, yaitu 500 poin (Mullis et al., 2015).

Ketidakmampuan siswa untuk mencapai standar yang ditetapkan oleh asesmen global dapat

disebabkan oleh minimnya pengembangan soal serupa di sekolah. Proses pembelajaran di sekolah masih belum terfokus pada pengembangan instrumen yang sesuai dengan tolak ukur internasional seperti TIMSS.

Penerapan penilaian numerasi berbasis TIMSS di sekolah dasar diharapkan dapat meningkatkan kualitas pembelajaran dan evaluasi numerasi. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan instrumen tes numerasi berbasis TIMSS yang valid dan reliabel, serta mampu mengukur kemampuan numerasi siswa secara komprehensif sehingga dapat memberikan kontribusi bagi pengembangan metode evaluasi yang lebih efektif dan sesuai dengan kebutuhan siswa. Instrumen tes numerasi yang dikembangkan dianalisis validitas, reliabilitas, karakteristik butir soal, kepraktisan instrumen, serta profil numerasi siswa berdasarkan hasil uji coba tes tersebut.

(Arikunto, 2013) menjelaskan bahwa tes merupakan kumpulan pertanyaan atau alat ukur yang digunakan untuk menilai keterampilan, pengetahuan,

kecerdasan, dan bakat individu atau kelompok. (Arikunto, 2008) juga menyatakan bahwa kualitas sebuah tes dapat dinilai berdasarkan lima kriteria, yaitu: (1) Validitas, yang mengukur sejauh mana tes secara akurat menilai hasil belajar dan

menghasilkan data yang sesuai; (2) Reliabilitas, yang menunjukkan kestabilan hasil tes tanpa dipengaruhi oleh faktor eksternal; (3) Objektivitas, yang memastikan bahwa hasil tes tidak dipengaruhi oleh unsur subjektivitas atau bias; (4) Praktikabilitas, yang berfokus pada kemudahan dalam melaksanakan, memeriksa, dan memberikan petunjuk sehingga bisa diterapkan oleh orang lain; (5) Ekonomis, yang berarti tes dapat dilaksanakan dengan biaya rendah dan efisien dalam penggunaan waktu.

TIMSS matematika diadakan untuk kelas IV dan kelas VII. Kerangka penilaian matematika TIMSS disusun dalam dua dimensi:

1. Dimensi konten, menentukan domain materi pelajaran yang akan dinilai
2. Dimensi kognitif, merinci proses berpikir yang akan dinilai

Pada TIMSS 2023 *Assessment Frameworks*, dijelaskan pembagian

konten konten dan presentase masing-masing domain konten seperti pada Tabel 1 berikut:

**Tabel 1 Presentase masing-masing domain dalam domain ko TIMSS Matematika kelas IV**

Domain Konten	Presentase
Bilangan	50%
Pengukuran dan Geometri	30%
Data	20%

Dalam asesmen TIMSS, terdapat 3 domain kognitif, yaitu pengetahuan 40%, penerapan 40%, dan penalaran 20%(Mullis et al., 2023).

## **B. Metode Penelitian**

Desain penelitian yang digunakan dalam penelitian ini merupakan desain penelitian dan pengembangan atau R&D (*Research and Development*). Menurut (Sugiyono, 2020), penelitian dan pengembangan merupakan penelitian yang menghasilkan produk dan menguji keefektifan produk tersebut.

Model pengembangan yang digunakan dalam penelitian ini adalah model Four-D (4D) yang dikembangkan oleh Sivasailam Thiagarajan, Dorothy S. Semmel, dan Melvyn I. Semmel. Model ini terdiri dari empat tahap utama, yaitu

pendefinisian (*define*), perancangan (*design*), pengembangan (*develop*), dan penyebaran (*disseminate*).

Dalam tahap pendefinisian, analisis kebutuhan pengembangan dan syarat-syarat pengembangan produk yang sesuai dengan kebutuhan pengguna dilakukan. Tahap pendefinisian penelitian ini mencakup serangkaian kegiatan untuk mengidentifikasi kebutuhan pengembangan instrumen penilaian numerasi.

Pada tahap perancangan, peneliti membuat produk awal, yang disebut sebagai prototype, atau rancangan produk yang disesuaikan dengan kerangka isi yang dihasilkan dari analisis kurikulum dan materi (Slamet, 2022). Dalam penelitian ini, tahap perancangan difokuskan pada pengembangan instrumen tes numerasi. Kegiatan utama yang dilakukan adalah menyusun kisi-kisi soal dan menulis soal-soal tes numerasi berdasarkan hasil analisis data pada tahap sebelumnya.

Tahap pengembangan dibagi menjadi dua tahap yaitu; *expert appraisal* dan *development testing* (Thiagarajan et al., 1974). *Expert appraisal* adalah teknik validasi atau menilai kelayakan rancangan produk, Dalam kegiatan ini, ahli di bidang

tersebut melakukan evaluasi. Produk yang telah dibuat diperbaiki dengan saran yang diberikan. Sementara *development testing* adalah kegiatan uji coba rancangan produk pada sasaran atau subjek penelitian.

Tahap penyebaran hasil penelitian ini dilakukan melalui beberapa media, yaitu publikasi di jurnal ilmiah bereputasi dan pemaparan dalam seminar hasil tesis. Tujuan dari tahap penyebaran ini adalah supaya produk dapat dimanfaatkan oleh orang lain.

Sumber data pada penelitian ini beragam. Data kualitatif diperoleh melalui wawancara dengan guru, siswa, dan observasi kelas. Data kuantitatif diperoleh dari hasil validasi 5 orang ahli yaitu ahli evaluasi, 2 ahli matematika (dosen matematika dan guru matematika), dan 2 ahli pendidikan dasar (dosen PGSD dan guru kelas SD). Sumber data lainnya diperoleh dari hasil uji coba tes yang diikuti 109 siswa kelas IV di lima sekolah dasar di Kecamatan Ponggok pada tahun ajaran 2024/2025. Uji coba skala kecil dilaksanakan di SD Negeri Ponggok 01 dengan jumlah peserta tes 19 siswa, sedangkan uji coba skala besar diikuti oleh 90 siswa dengan rincian 24 siswa SD Negeri Ponggok 03, 21 siswa SD Negeri

Ponggok 04, 29 siswa SD Candirejo 1 dan 16 siswa SD Negeri Bendo 2.

### C. Hasil Penelitian dan Pembahasan

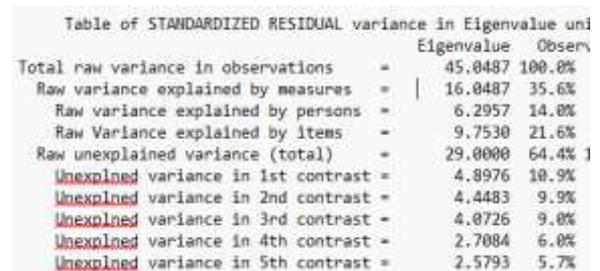
#### 1. Validitas Instrumen Tes

Dalam proses penilaian validitas instrumen, terdapat beberapa jenis uji validitas yang digunakan. Pertama, validitas isi, yang melibatkan para ahli dalam mengevaluasi kesesuaian butir-butir tes yang disusun. Kedua, validitas konstruk, yang menelaah sejauh mana instrumen mencerminkan konsep teoritis yang melandasinya. Ketiga, validitas kriteria, yang juga dikenal sebagai validitas empiris, yang mengkaji hubungan antara hasil tes dengan ukuran atau kriteria eksternal yang relevan.

Uji validitas isi terhadap instrumen tes dilakukan dengan menganalisis penilaian para ahli terhadap instrumen menggunakan rumus Aiken's V yang dibantu dengan aplikasi Microsoft Excel. Berdasarkan hasil uji validitas isi, keseluruhan butir soal atau 30 butir soal memiliki validitas > 0.80 sehingga termasuk dalam kategori sangat valid. Instrumen dinyatakan layak digunakan sebagai alat ukur untuk mengukur kemampuan numerasi siswa kelas IV sekolah dasar.

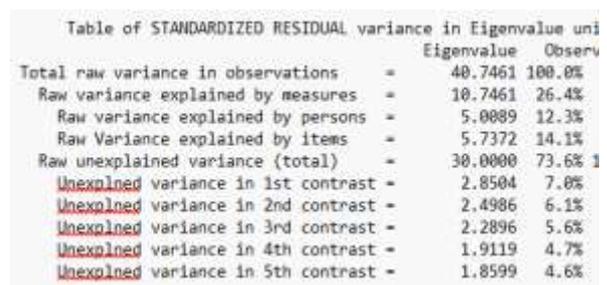
Dalam pendekatan IRT, analisis konstruk dari perangkat tes dapat dilakukan melalui pengujian terhadap asumsi unidimensionalitas (Curtis, 2004). Menurut (Sumintono & Widhiarso, 2015), asumsi unidimensi ini merupakan salah satu ukuran penting dalam analisis instrumen, yang digunakan untuk menentukan apakah instrumen yang dikembangkan telah memiliki validitas konstruk yang memadai.

**Gambar 1 Item Dimensionality Uji Skala Kecil**



	Eigenvalue	Observ
Total raw variance in observations	45.0487	100.0%
Raw variance explained by measures	16.0487	35.6%
Raw variance explained by persons	6.2957	14.0%
Raw Variance explained by items	9.7530	21.6%
Raw unexplained variance (total)	29.0000	64.4%
Unexplned variance in 1st contrast	4.8976	10.9%
Unexplned variance in 2nd contrast	4.4483	9.9%
Unexplned variance in 3rd contrast	4.0726	9.0%
Unexplned variance in 4th contrast	2.7084	6.0%
Unexplned variance in 5th contrast	2.5793	5.7%

**Gambar 2 Item Dimensionality Uji Skala Besar**



	Eigenvalue	Observ
Total raw variance in observations	40.7461	100.0%
Raw variance explained by measures	10.7461	26.4%
Raw variance explained by persons	5.0089	12.3%
Raw Variance explained by items	5.7372	14.1%
Raw unexplained variance (total)	30.0000	73.6%
Unexplned variance in 1st contrast	2.8504	7.0%
Unexplned variance in 2nd contrast	2.4986	6.1%
Unexplned variance in 3rd contrast	2.2896	5.6%
Unexplned variance in 4th contrast	1.9119	4.7%
Unexplned variance in 5th contrast	1.8599	4.6%

Berdasarkan Gambar 1 dan 2, diketahui nilai *Raw Variance Explained by Measure* pada uji skala kecil sebesar 35.6 % dan 26.4% pada uji skala besar. Hal ini menunjukkan bahwa instrumen yang dikembangkan telah mampu mengukur konstruk yang dimaksud.

(Misbach & Sumintono, 2014) menjelaskan bahwa untuk memenuhi asumsi unidimensionalitas, diperlukan minimal nilai sebesar 20%, sedangkan untuk unidimensi diatas 40% dinyatakan sangat baik.

Pengujian validitas instrumen tes berdasarkan besarnya validitas setiap butir dikenal dengan istilah *concurrent validity* atau validitas empiris. Pembuktian validitas ini dilakukan melalui analisis model Rasch dengan meninjau tingkat kesesuaian butir soal. penentuan *item fit* dilakukan berdasarkan kriteria *item fit*. Tabel 2 berikut menyajikan pengelompokan butir soal yang memenuhi kriteria *item fit*.

**Tabel 2 Pengelompokan Kriteria Butir Tes Berdasarkan *Item Fit***

<b>Kategori Kriteria</b>	<b>Kriteria</b>	<b>Jumlah soal</b>	<b>Nomer Soal</b>
Satu Kriteria	MNSQ	1 soal	19
Dua Kriteria	MNSQ dan ZSTD	17 soal	5, 7, 8, 9, 11, 13, 14, 15, 17, 20, 21, 22, 25, 27, 28, 29, 30
Tiga Kriteria	MNSQ, ZSTD dan PT Measure Corr	12 soal	1, 2, 3, 4, 6, 10, 12, 16, 18, 23, 24, dan 26

Menurut (Palimbong et al., 2018), validitas butir masih dapat diterima jika memenuhi dua dari tiga kriteria dan soal harus diperbaiki ketika hanya mempunyai satu kriteria yang sesuai. (Linacre, 2010) juga menekankan pentingnya fokus pada MNSQ dalam model Rasch. Berdasarkan tiga kriteria fit, 17 butir soal telah memenuhi tiga kriteria, 12 butir soal memenuhi dua kriteria (MNSQ dan *Pt Mean Corr*), dan satu soal memenuhi satu kriteria (MNSQ) namun tetap dapat dipertahankan dengan perbaikan untuk meningkatkan kualitas soal.

## 2. Reliabilitas Instrumen Tes

Pengujian reliabilitas instrumen tes numerasi berbasis TIMSS dilaksanakan melalui dua tahapan. Proses pengujian ini melibatkan penilaian terhadap kesesuaian pendapat para ahli serta pengukuran tingkat reliabilitas berdasarkan butir soal dan tanggapan responden. Uji reliabilitas oleh para ahli dilakukan dengan mengacu pada nilai *Cronbach's Alpha* serta pendekatan *Interclass Correlation Coefficient (ICC)* dengan bantuan aplikasi SPSS. Selain menggunakan *software* SPSS, untuk mengetahui reliabilitas instrumen tes dapat dilakukan uji reliabilitas

menggunakan analisis model Rasch melalui *software* Winsteps.

**Gambar 3 Hasil Uji Reliabilitas Cronbach's Alpha**

Reliability Statistics				
Cronbach's Alpha		N of Items		
.720		5		

Intraclass Correlation Coefficient				
	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		Value
		Lower Bound	Upper Bound	
Single Measures	.340 <sup>a</sup>	.186	.525	3.570
Average Measures	.720 <sup>c</sup>	.534	.847	3.570

Berdasarkan Gambar 3, diperoleh nilai *Cronbach's Alpha* sebesar 0,72 dari penilaian para ahli terhadap instrumen. Suatu instrumen dinyatakan memiliki keandalan yang baik apabila nilai koefisien *Cronbach's Alpha* mencapai atau melebihi angka 0,6. Selain itu, uji *Interclass Correlation Coefficient* (ICC) nilai koefisien sebesar 0,72. Jika nilai ICC berada dalam rentang  $0,5 < r \leq 0,75$ , maka reliabilitasnya tergolong tinggi.

Hasil analisis model Rasch menunjukkan bahwa nilai *person reliability* sebesar 0,79, yang mengindikasikan bahwa konsistensi responden dalam menjawab soal tergolong cukup. Sementara itu, nilai *item reliability* sebesar 0,93 menunjukkan bahwa butir-butir soal yang dikembangkan memiliki tingkat konsistensi yang baik dan tergolong dalam kategori reliabilitas tinggi.

### 3. Karakteristik Butir Tes

#### A. Tingkat Kesukaran

Distribusi tingkat kesulitan setiap butir soal disajikan pada Tabel 3 berikut;

**Tabel 3 Hasil Pengelompokan Tingkat Kesukaran Butir Tes Pada Uji Skala Besar (SD = 0,96)**

Interval	Kriteria	Butir Soal	Jumlah	Presentase
$0,0 \leq$ <i>Measure logit</i> $> +$ <i>0,96</i>	Sangat Sulit	30,	4	13,3%
		12,		
$0,0 \leq$ <i>Measure logit</i> $\leq +$ <i>0,96</i>	Sulit	8,	12	40%
		19,		
$-0,96 \leq$ <i>Measure logit</i> $\leq$ <i>0,0</i>	Mudah	25,	9	30%
		20,		
$0,0 \leq$ <i>Measure logit</i> $< -$ <i>0,96</i>	Sangat Mudah	18,	5	16,7%
		9,		
		11,		
		17,		
		13		

Pada uji skala besar diperoleh standar deviasi sebesar 0,96. Butir nomor 30 merupakan butir soal paling sulit yang memiliki nilai logit sebesar 1,98 dengan jawaban benar sebanyak 14 dari total 90 responden. Sebaliknya butir nomor 13 merupakan butir soal paling mudah dengan nilai logit sebesar -2,17 yang

mendapat jawaban benar sebanyak 78 dari 90 responden. Dalam pengelompokan tingkat kesukaran diketahui terdapat soal dengan kategori paling sulit sebanyak 4 soal, sulit 12 soal, mudah 9 soal dan sangat mudah 5 soal atau jika dipresentasikan soal sangat sulit sebesar 13,3 %, sulit 40%, mudah 30% dan sangat mudah 16,7%.

**B. Estimasi Kesalahan Butir (Daya Beda)**

Hasil analisis daya beda uji skala besar, diketahui semua butir soal ada pada kriteria daya beda yang baik. Sejumlah 30 soal memiliki nilai Standar Error dibawah 0,5. Dengan demikian semua butir soal dapat dinyatakan sebagai butir soal yang teliti dalam mengukur kemampuan numerasi siswa.

**C. Deteksi Item Bias (DIF)**

Berdasarkan hasil analisis DIF, dari 30 butir instrumen tes numerasi berbasis TIMSS yang telah diujikan, seluruh butir soal baik pada uji skala kecil maupun uji skala besar menunjukkan nilai probabilitas (PROB) > 0,05. Hal ini mengindikasikan bahwa tidak ada satu pun butir yang mengandung DIF, sehingga seluruh item dapat dinyatakan bebas dari bias yang dapat merugikan atau menguntungkan responden tertentu.

**4. Kepraktisan Instrumen**

Hasil penilaian kepraktisan instrumen tes numerasi berbasis TIMSS disajikan dalam Tabel 4 berikut.

**Tabel 4 Hasil Penilaian Kepraktisan**

PE NILAI	Aspek 1		Aspek 2		Aspek 3		Aspek 4		Aspek 5	
	1	2	3	4	5	6	7	8	9	10
P1	3	4	3	3	4	4	4	4	3	4
P2	4	4	4	4	4	4	4	4	4	4
P3	4	3	4	3	4	3	4	3	3	4
P4	4	3	3	4	4	3	4	4	3	4
P5	4	3	4	4	4	3	4	4	3	4
Total	19	17	18	18	20	17	20	19	16	20
Skor						184				
						92 %				

**Instrumen Tes Numerasi Berbasis TIMSS**

Berdasarkan tabel hasil penilaian kepraktisan instrumen, jumlah skor penilaian kepraktisan yang didapatkan sebesar 184, sedangkan jumlah skor ideal seluruh item adalah 200. Sehingga didapatkan nilai kepraktisan sebesar 92%. Merujuk pada kriteria kepraktisan, nilai kepraktisan sebesar 92% termasuk dalam kriteria sangat praktis.

**5. Profil Numerasi Siswa**

Hasil analisis kemampuan responden menggunakan analisis Rasch menunjukkan dari 90 responden yang mengikuti uji coba tes, sebanyak 12 responden (13,3%) responden memiliki kemampuan tinggi, 64 responden (71,1%) berada pada tingkat kemampuan sedang, dan sisanya 14 responden (15,65%) termasuk dalam

kategori kemampuan rendah seperti tergambar pada Gambar 4 berikut:

**Gambar 4 Sebaran Kemampuan Responden**



**E. Kesimpulan**

Hasil analisis terhadap instrumen tes numerasi berbasis TIMSS yang dilakukan menunjukkan bahwa instrumen tes numerasi yang dikembangkan merupakan instrumen yang valid, reliabel, dan praktis. Instrumen tes numerasi berbasis TIMSS dapat mengukur kemampuan numerasi siswa kelas IV dengan efektif serta mengidentifikasi kelemahan dan kelebihan siswa sehingga dapat

menentukan strategi pembelajaran yang lebih sesuai dan tepat sasaran. Kemudian, peneliti berharap dapat membuat instrumen serupa untuk berbagai jenjang kelas pada penelitian selanjutnya agar dapat memperluas konteks penggunaan instrumen.

**DAFTAR PUSTAKA**

Adinda, D. W., Nurhasanah, & Oktavianti, I. (2022). Profil Kemampuan Numerasi Dasar Siswa Sekolah Dasar Di SDN Mentokan. *Jurnal Ilmiah Profesi Pendidikan*, 7(3), 1066–1070. <https://doi.org/10.29303/jipp.v7i3.700>

Arikunto. (2008). *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara.

Arikunto, S. (2013). *Dasar-dasar Evaluasi Pendidikan* (Jakarta). Bumi Aksara.

Curtis, D. D. (2004). Person Misfit in Attitude Surveys: Influences, Impacts and Implications. *International Education Journal*, 5(2), 125–143.

Dougherty, C. (2003). Numeracy, literacy and earnings: evidence from the National Longitudinal Survey of Youth. *Economics of Education Review*, 22(5), 511–521. [https://doi.org/https://doi.org/10.1016/S0272-7757\(03\)00040-2](https://doi.org/https://doi.org/10.1016/S0272-7757(03)00040-2)

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., &

- Beilock, S. L. (2017). The Math Anxiety-Performance Link: A Global Phenomenon. *Current Directions in Psychological Science*, 26(1), 52–58.  
<https://doi.org/10.1177/0963721416672463>
- Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28), 11267–11271.  
<https://doi.org/10.1073/pnas.1220568110>
- Gross, J., Çekiç, O., Hossler, D., & Hillman, N. (2009). What Matters in Student Loan Default: A Review of the Research Literature. *Journal of Student Financial Aid*, 39, 19–29.  
<https://doi.org/10.55504/0884-9153.1032>
- Kemendikbud. (2023). *Rapor Pendidikan Indonesia 2023*.
- Linacre, J. M. (2010). Predicting responses from rasch measures. *Journal of Applied Measurement*. *Journal of Applied Measurement*, 11(1), 1–10.
- Maulidina, A. P. (2019). Profil Kemampuan Numerasi Siswa Sekolah Dasar Berkemampuan Tinggi Dalam Memecahkan Masalah Matematika. *Jurnal Bidang Pendidikan Dasar*, 3(2), 61–66.  
<https://doi.org/10.21067/jbpd.v3i2.3408>
- Misbach, I. H., & Sumintono, B. (2014, May 24). “Pengembangan dan Validasi Instrumen Persepsi Siswa Terhadap Karakter Moral Guru di Indonesia dengan Model Rasch.” *Seminar Nasional Psikometri: “Pengembangan Instrumen Penilaian Karakter Yang Valid. Di Hotel Lorin Solo*.
- Mullis, I. V. S., Davier, M. V., & Martin, M. O. (2023). *TIMSS 2023 Assessment Frameworks*. IEA TIMSS & PIRLS International Study Center.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2015). *TIMSS 2015 International Results in Mathematics*.
- Palimbong, J., Mujasam, M., & Allo, A. Y. T. (2018). Item Analysis Using Rasch Model in Semester Final Exam Evaluation Study Subject in Physics Class X TKJ SMK Negeri 2 Manokwari. *Kasuari: Physics Education Journal (KPEJ)*, 1(1), 43–51.
- Prastyo, H. (2020). Kemampuan Matematika Siswa Indonesia Berdasarkan TIMSS. *Jurnal Padeagogik*, 3(2), 111–117.  
<https://doi.org/10.35974/jpd.v3i2.2367>
- Slamet, F. A. (2022). *Model Penelitian Pengembangan (R n D)*. Institut Agama Islam Sunan Kalijogo Malang.
- Sugiyono. (2020). *Metodologi Penelitian Kuantitatif, Kualitatif dan R & D*. Alfabeta.

Sumintono, B., & Widhiarso, W. (2015).  
*Aplikasi Pemodelan Rasch Pada  
Assesment Pendidikan*. Trim  
Komunikata Publishing House.

Tenny, Nisa, A. K., & Murtaplah.  
(2021a). *Pengembangan Literasi  
dan Numerasi dalam Proses  
Belajar dan Mengajar*.

Tenny, Nisa, A. K., & Murtaplah.  
(2021b). *Pengembangan Literasi  
dan Numerasi dalam Proses  
Belajar dan Mengajar*. Direktorat  
Sekolah Menengah Atas.  
[http://repositori.kemdikbud.go.id/id/  
eprint/29935](http://repositori.kemdikbud.go.id/id/eprint/29935)

Thiagarajan, S., Semmel, D. S., &  
Semmel, M. I. (1974). *Instructional  
development for Training Teachers  
of Exceptional Children: A  
Sourcebook*.