

**KUALITAS BUTIR SOAL SUMATIF TENGAH SEMESTER (STS) BUATAN GURU
PADA MATA PELAJARAN BAHASA INDONESIA KELAS III SDN 11
PANGKALPINANG**

^{1*}Asyraf Suryadin, ²Della Fitria

^{1,2}Universitas Muhammadiyah Bangka Belitung

email korespondensi: ^{1*}asyraf.suryadin@unmuhbabel.ac.id, ²dellafitria367@gmail.com

ABSTRACT

This study aims to analyze the Summative Middle Semester multiple choice question grain of Indonesian Language grade III elementary school Public 11 Pangkalpinang based on the index of difficulty, differential power, confounding effectiveness, validity and reliability. The method of data collection in this study was the descriptive quantitative approach documentation method. Based on the results of the study found the difficulty level was not good but had sufficient differential power, question tampering was effective or already good, question validity was categorized as high and reliability scores were still low with unreliable status. However some questions are revised or need to be replaced with other question grain especially the questions which are still poor. The results of this study are expected to reference teachers in improving the preparation of subsequent Mid-Semester Summative (STS) questions.

Keywords: Question Item Analysis, Mid-Semester Summative, Indonesian Language

ABSTRAK

Penelitian ini bertujuan untuk menganalisis butir soal pilihan ganda Sumatif Tengah Semester (STS) materi Bahasa Indonesia kelas III SD Negeri 11 Pangkalpinang berdasarkan indeks kesukaran, daya beda, efektifitas pengecoh, validitas dan reabilitas. Metode pengumpulan data dalam penelitian ini adalah metode dokumentasi pendekatan kuantitatif deskriptif. Berdasarkan hasil penelitian ditemukan tingkat kesukaran tidak baik tetapi memiliki daya beda yang cukup, pengecoh soal efektif atau sudah baik, validitas soal dikategorikan tinggi dan nilai reliabilitas yang masih rendah dengan status tidak reliabel. Namun ada beberapa soal direvisi atau perlu diganti dengan butir soal yang lain terutama soal-soal yang masih kurang baik. Hasil penelitian ini diharapkan dapat menjadi rujukan guru dalam memperbaiki penyusunan soal Sumatif Tengah Semester (STS) selanjutnya.

Kata kunci: Analisis Butir Soal, Sumatif Tengah Semester, Bahasa Indonesia

A. PENDAHULUAN

Kegiatan penilaian selalu disamaartikan dengan istilah kegiatan evaluasi. Evaluasi merupakan bagian dari proses dan secara keseluruhan tidak dapat dipisahkan dari kegiatan pembelajaran. Kegiatan evaluasi telah diatur dalam UndangUndang Republik Indonesia No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional Bab XVI Pasal 58 Ayat 1, menyatakan bahwa "Evaluasi hasil belajar peserta didik dilakukan oleh pendidik untuk memantau proses, kemajuan, dan perbaikan hasil belajar peserta didik secara berkesinambungan." Oleh karena itu, evaluasi hasil belajar bertujuan untuk menilai pencapaian kompetensi dan memperbaiki proses pembelajaran serta pedoman penyusunan laporan kemajuan hasil belajar siswa (T. Kurniawan, 2015). Permasalahannya apakah guru telah menyusun evaluasi hasil belajar dengan benar?

Untuk mengetahui hasil belajar siswa diawali dengan tes. Tes salah satu alat yang digunakan sebagai sarana untuk penilaian hasil belajar. Menurut (Sudijono, 2017) tes adalah cara dalam mengukur dan menilai di bidang pendidikan dalam bentuk pemberian tugas atau serangkaian tugas yang harus dikerjakan oleh siswa, sehingga dapat diketahui nilai prestasi siswa. Nilai tersebut kemudian dibandingkan dengan nilai-nilai yang dicapai oleh testee lainnya, atau dibandingkan dengan nilai standar tertentu. Oleh karena itu, untuk mengetahui kualitas butir soal yang digunakan untuk tes perlu dilakukan analisis butir soal (Magdalena, Fauziah, et al., 2021). Selanjutnya, Arikunto (2018) menyatakan bahwa tes dikatakan baik apabila memenuhi syarat validitas, reliabilitas, objektivitas,

praktikabilitas, dan ekonomis. Permasalahannya tes yang dijadikan bahan evaluasi sering kali tidak dilakukan analisis kualitas butir soal oleh guru, sehingga belum diketahui kualitas soal yang telah disusunnya. Oleh karena itu, untuk mengetahui kualitas butir soal yang digunakan untuk tes perlu dilakukan analisis butir soal (Magdalena, Anggraini, et al., 2021).

Kegiatan menganalisis butir soal merupakan kegiatan yang harus dilakukan guru untuk meningkatkan mutu soal yang telah ditulis. Kegiatan ini merupakan proses pengumpulan, peringkasan dan penggunaan informasi dari jawaban siswa untuk membuat keputusan tentang setiap penilaian (R. Y. Kurniawan et al., 2017). Analisis ini memberikan informasi tentang kualitas butir soalyang dibuat oleh guru dilihat dari tingkat kesukaran, daya beda, efektivitas pengecoh, validitas dan reabilitas (Farhan et al., 2021). Tujuan menganalisis soal buatan guru ini adalah untuk mengkaji dan menelaah setiap butir soal agar diperoleh soal yang bermutu sebelum soal digunakan oleh guru kepada siswa. Selain itu, tujuan analisis butir soal juga untuk membantu meningkatkan tes melalui revisi atau membuang soal yang tidak efektif, serta untuk mengetahui informasi diagnostik pada siswa apakah mereka sudah atau belum memahami materi yang telah diajarkan (Destiniar et al., 2018). Hasil belajar siswa juga dipengaruhi oleh kemampuan guru membuat soal. Baik dan buruknya hasil yang didapatkan bergantung pada kualitas soal yang dibuat oleh masing-masing guru (Rahmasari & Ismiyati, 2016).

Kurangnya perhatian guru dalam membuat soal-soal yang baik, maka

soal buatan guru hanya mempunyai daya jangkau pakai yang terbatas. Hasil atau skor yang dicapai siswa juga terbatas, dalam arti hanya diperbandingkan dengan teman-teman sekelompoknya yang satu sekolah. Perbandingan antara siswa dari satu sekolah dengan sekolah yang lain kurang tepat karena mungkin sekali alat ukur dan cara penafsiran yang dilakukan guru yang mengetes tidak sama. Jadi, apa yang disusun guru pada waktu itu, kemudian yang dievaluasi bahkan mungkin untuk berkali-kali. Kegiatan analisis dan revisi butir-butir soal jarang dilakukan. Itulah sebabnya materi, kontruksi soal bahasa, validitas, reliabilitas dan analisis butir soal yang terdiri dari tingkat kesukaran, daya pembeda dan pengecoh soal buatan guru sering dikatakan rendah, atau sebenarnya yang tepat adalah tidak diketahui secara pasti karena memang jarang dilakukan pengujian terhadap alat evaluasi (Suryadin A dkk.,2020)

Berdasarkan latar belakang tersebut, akan dianalisis kualitas butir soal yang dibuat oleh guru SD Negeri 11 Pangkalpinang. Tes tertulis telah dilaksanakan pada 31 Mei 2024, di kelas III SD Negeri 11 Pangkalpinang. Tes tersebut berupa soal Sumatif Tengah Semester (STS) materi Bahasa Indonesia yang dibuat oleh guru. Soal STS yang diujikan yaitu soal pilihan ganda dan isian. Namun, analisis butir soal hanya berfokus pada soal pilihan ganda. Kualitas butir soal yang dianalisis meliputi: indeks kesukaran, daya beda, efektivitas pengecoh, validitas dan reabilitas. Tujuan penelitian ini adalah untuk memberikan gambaran yang nyata tentang evaluasi terhadap soal Sumatif Tengah Semester yang telah disusun dan dijadikan masukan pada penyusunan soal selanjutnya sehingga

evaluasi yang dikerjakan oleh siswa benar-benar memenuhi kualitas butir soal.

B. METODE PENELITIAN

Metode pengumpulan data dalam penelitian ini adalah metode dokumentasi. Dokumentasi yaitu mencari data mengenai hal-hal atau variabel yang berupa catatan, transkrip, buku, surat kabar, majalah, prasasti, notulen rapat, legger, agenda, dan sebagainya (Arikunto, 2010). Dengan demikian dokumen diperoleh dengan menggandakan soal yang telah dibuat oleh guru, kunci jawaban dan lembar jawab siswa. Data yang telah didapat dianalisis tingkat kesukaran, daya pembeda, pengecoh, validitas, dan reliabilitas. Penelitian ini berbentuk analisis deskriptif dengan pendekatan kuantitatif. Analisis deskriptif adalah analisis yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul (Tilaar & Hasriyanti, 2019). Selanjutnya, akan dilengkapi sebagai bahan pendukung penelitian secara kuantitatif. Penelitian kuantitatif adalah suatu proses menemukan pengetahuan yang menggunakan data berupa angka sebagai alat menganalisis keterangan mengenai apa yang ingin kita ketahui (Djollong, 2014). Sedangkan subjek penelitian ini adalah 30 siswa kelas III SD di Negeri 11 Pangkalpinang.

C. HASIL DAN PEMBAHASAN

1. Indeks Kesukaran

Tingkat kesukaran soal merupakan pengukuran seberapa besar derajat kesukaran soal tersebut. Suatu soal dikatakan baik, apabila memiliki tingkat kesukaran soal yang seimbang dalam

artian soal tersebut tidak terlalu mudah atau terlalu sukar. Pembagian tingkat kesukaran ini dapat dibagi menjadi tiga tingkatan, yaitu sukar, sedang dan mudah. Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha untuk memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa. Bilangan yang menunjukkan sukar mudahnya soal disebut indeks kesukaran (*difficulty index*) yang diberi simbol dengan huruf P. Besarnya indeks

kesukaran soal antara 0,00 sampai 1,0. Semakin besar indeks kesukaran soal maka semakin mudah soal tersebut dan sebaliknya makin kecil indeks yang diperoleh, makin sulit soal tersebut. Dengan kriteria, soal dengan P 0,00 – 0,30 adalah soal sukar; soal dengan P 0,31 – 0,70 adalah soal sedang; soal dengan P 0,71 – 1,00 adalah soal mudah (Fatimah & Alfath, 2019).

Tingkat Kesukaran soal STS Bahasa Indonesia kelas III SD Negeri 11 Pangkajene dapat terlihat pada tabel 1 berikut.

Tabel 1. Gambaran Tingkat Kesukaran Butir Soal Pilihan Ganda

No soal	1	2	3	4	5	6	7	8	9	10
Jumlah Jawaban Benar	28	21	22	15	16	24	27	13	12	21
Indeks Kesukaran	0,93	0,70	0,73	0,50	0,53	0,80	0,90	0,43	0,40	0,70
Keterangan	Mudah	Sedang	Mudah	Sedang	Sedang	Mudah	Mudah	Sedang	Sedang	Sedang

Berdasarkan hasil analisis soal pilihan ganda menunjukan bahwa soal yang memiliki tingkat kesukaran dengan kategori sukar berjumlah 0 atau sebesar 0%, soal yang memiliki tingkat kesukaran sedang berjumlah 6 atau sebesar 60% dan dan butir soal yang memiliki tingkat kesukaran mudah berjumlah 4 soal atau sebesar 40%. Berdasarkan hasil yang didapatkan dapat disimpulkan bahwa tingkat kesukaran soal buruk (tidak baik) dalam evaluasi dan tidak dapat digunakan dalam evaluasi selanjutnya dikarenakan tidak memiliki soal yang sukar.

2. Daya Beda

Daya pembeda adalah kemampuan sesuatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah). Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi, disingkat D (d besar). Seperti halnya indeks kesukaran, indeks diskriminasi (daya pembeda) ini berkisar antara 0,00 sampai 1,00. Hanya bedanya, indeks kesukaran tidak mengenal tanda negatif (-), tetapi pada indeks diskriminasi ada tanda negatif (-). Tanda negatif pada indeks diskriminasi digunakan jika sesuatu soal “terbaik”

menunjukkan kualitas teste. Yaitu anak

pandai disebut bodoh dan anak bodoh disebut pandai (Suryadin,A & Haiyudi, 2023).

Indonesia kelas III SD Negeri 11 Pangkajene dapat terlihat pada tabel 2 berikut.

Daya beda soal STS Bahasa

Tabel 2. Daya Pembeda Butir Soal Pilihan Ganda

No soal	1	2	3	4	5	6	7	8	9	10
Daya Beda	0,00	0,33	0,40	0,33	0,53	0,27	0,20	0,20	0,13	0,33
Keterangan	Jelek	Cukup	Cukup	Cukup	Baik	Cukup	Jelek	Jelek	Jelek	Cukup

Hasil analisis daya pembeda menunjukkan bahwa soal yang memiliki daya pembeda jelek berjumlah 4 soal atau sebesar 40 %, soal yang memiliki daya pembeda cukup berjumlah 5 soal atau sebesar 50 %, soal yang memiliki daya pembeda baik berjumlah 1 soal atau sebesar 10 % dan soal yang memiliki daya pembedanya baik sekali berjumlah 0 atau sebesar 0 %. Dapat disimpulkan bahwa dikelas III SD Negeri 11 Pangkajene yang mendominasi yaitu siswa yang pandai atau berkemampuan tinggi.

3. Efektivitas Pengecoh (Distraktor)

Indeks pengecoh (*distractor*) adalah suatu analisis pola jawaban pada soal pilihan ganda untuk melihat bagaimana pilihan jawaban yang dipilih

oleh siswa atau responden (Suryadin, A & Haiyudi, 2023). Jika pilihan jawaban pada setiap item sekurang-kurangnya dipilih oleh siswa sebesar 5% maka distraktor dinyatakan sudah menjalankan fungsinya dengan benar (Mochammad Noor Akhmadi, 2021). Pedoman yang digunakan dalam langkah yang perlu dilakukan untuk menganalisis pengecoh ialah dengan menghitung banyaknya siswa yang memilih pilihan jawaban A, B, C, dan D, atau tidak memilih jawaban sama sekali yang disebut dengan *ommit* disingkat OM (Suryadin, A & Haiyudi, 2023).

Efektifitas pengecoh soal STS Bahasa Indonesia kelas III SD Negeri 11 Pangkajene dapat terlihat pada tabel 3 berikut.

Tabel 3. Pengecoh (Distraktor) Butir Soal Pilihan Ganda

Keterangan										
No Soal	1	2	3	4	5	6	7	8	9	10
A	DITOLAK	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA
B	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA
C	DITERIMA	DITERIMA	DITERIMA	DITERIMA	DITERIMA	REVISI	DITOLAK	DITERIMA	DITERIMA	DITERIMA

Berdasarkan hasil perhitungan, siswa memilih jawaban A yang diterima berjumlah 9 soal, jawaban B yang diterima berjumlah 10 soal, jawaban C yang diterima berjumlah 8 soal. Selanjutnya siswa memilih jawaban A yang ditolak berjumlah 1, jawaban B yang ditolak berjumlah 0, jawaban C yang ditolak berjumlah 1 serta jawaban revisi hanya berjumlah 1 yaitu pada jawaban C. Jadi, dapat disimpulkan

bahwa soal STS tersebut memiliki efektivitas

pengecoh baik, karena hanya memiliki 2 soal (ditolak) berkategori tidak efektif. Persentase pengecoh butir soal pilihan ganda dapat dilihat pada tabel 4 berikut.

Tabel 4. Persentase Pengecoh Butir Soal Pilihan Ganda

Distraktor										
No soal	1	2	3	4	5	6	7	8	9	10
A	0%	23%	80%	27%	33%	17%	10%	40%	43%	23%
B	93%	70%	10%	50%	53%	80%	90%	13%	40%	70%
C	7%	7%	10%	23%	13%	3%	0%	47%	17%	7%
Jumlah	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

4. Validitas

Menurut Syamsurizal (2020) validitas soal merupakan derajat kesesuaian antara sesuatu soal dengan perangkat soal-soal lain. Validitas dapat diartikan sebagai ukuran berapa ketepatan alat ukur dalam melakukan fungsinya, sehingga akan mendapatkan hasil sesuai dengan apa yang hendak di ukur. Ada yang mengatakan bahwa valid itu sama dengan sah, sehingga validitas = kesahihan (Nizary & Kholik, 2021). Suatu alat ukur disebut memiliki validitas jika alat ukur tersebut isinya

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}}$$

layak mengukur objek yang seharusnya diukur dan sesuai dengan kriteria. Artinya ada kesesuaian antara alat ukur dengan fungsi pengukuran dan sasaran pengukuran. Validasi suatu instrumen evaluasi, tidak lain adalah derajat yang

menunjukkan di mana suatu tes mengukur apa yang hendak diukur. Dapat disimpulkan bahwa validitas adalah suatu cara atau alat yang digunakan untuk mengukur valid atau tidaknya suatu instrumen dan memiliki derajat ukur seperti tinggi, sedang, dan rendah. Validas bersifat lebih spesifik karena hanya dapat mengukur satu bidang (Suryadin,A & Haiyudi, 2023).

Untuk menghitung validitas pada soal buatan guru yang telah diuji di SD Negeri 11 Pangkalpinang kepada 30 siswa maka digunakan rumus *Product Moment Correlation*, dan rumus di bawah ini rumus untuk skor kasar sebagai berikut:

Keterangan:

r_{xy} = Koefisien korelasi antara instrumen X dan instrumen Y
X = Variabel X(instrumen X),

Y = Variabel Y(instrumen Y),
 N = Jumlah Peserta.

Namun jika kita hendak menghitung menggunakan excel, maka kita dapat menggunakan formula sebagaimana pada gambar berikut



Gambar 1. Formula Validitas

Dengan menggunakan formula validitas di atas maka dapat disimpulkan apabila r hitung lebih besar daripada r tabel dapat dinyatakan valid.

Sebaliknya, apabila r hitung lebih kecil daripada r tabel maka dapat dinyatakan tidak valid. Sementara untuk mengetahui r hitung sendiri digunakan rumus di excel yaitu =CORREL pada masing- masing skor kasar di setiap soal, dan untuk mengetahui r tabel dapat menggunakan tingkat signifikasi 0,05 untuk uji dua arah. Sementara itu, untuk derajat kebebasannya dapat diketahui dengan $N - 2$ yaitu total responden $30 - 2$ sama dengan 28, yang diketahui r tabel yang diperoleh sebesar 0,361. Berikut ini hasil perhitungan yang diperoleh dari soal uji coba yang telah dilakukan.

Validitas soal STS Bahasa Indonesia kelas III SD Negeri 11 Pangkajene dapat terlihat pada tabel 5 berikut.

Tabel 5. Validitas Butir Soal Pilihan Ganda

	No Soal	1	2	3	4	5	6	7	8	9	10
VALIDITAS	Jumlah Jawaban Benar	28	21	22	15	16	24	27	13	12	21
	r Tabel	0,361	0,361	0,361	0,361	0,361	0,361	0,361	0,361	0,361	0,361
	r Hitung	-0,147	0,446	0,480	0,485	0,544	0,411	0,415	0,288	0,103	0,446
	Kriteria	INVALID	VALID	VALID	VALID	VALID	VALID	VALID	INVALID	INVALID	VALID
	Tingkat Korelasi	Sangat Rendah	Cukup	Cukup	Cukup	Cukup	Cukup	Cukup	Rendah	Rendah	Cukup

Hasil analisis validitas butir soal diketahui bahwa soal yang valid berjumlah 7 soal atau sebesar 70 % sedangkan soal invalid berjumlah 3 atau sebesar 30%. Proporsi soal ditunjukkan soal valid yang lebih mendominasi, sehingga dapat dikategorikan soal divalidasi tinggi walaupun masih

terlihat adanya validasi yang rendah. Dalam beberapa kasus, keterbatasan waktu, sumber daya, dan tekanan administratif menyebabkan penyusunan evaluasi dilakukan secara terburu-buru, sehingga aspek validitas menjadi terabaikan. Permasalahan-permasalahan tersebut menunjukkan

bahwa isu validitas bukan hanya persoalan teknis, tetapi juga berkaitan dengan kesadaran, kompetensi, dan budaya evaluasi dalam lingkungan Pendidikan (Farida & Musyarofah, 2021).

5. Reliabilitas

Reliabilitas sendiri berkaitan dengan kepercayaan (Janna & Herianto, 2021). Sehingga reliabilitas merupakan tingkat kepercayaan atau keterandalan suatu alat ukur. Suatu tes itu dapat dipercaya apabila mempunyai taraf kepercayaan yang cukup tinggi. Syarat kepercayaan dalam tes merupakan suatu ketetapan. Tes yang dapat dipercaya adalah tes yang

mempunyai ketetapan hasil. Walaupun jika ada proses perubahan, itu tidak terlalu signifikan dan tidak mempengaruhi isi tes secara substansial. Tes yang reliabel adalah tes yang konsisten. Jika tes itu diulang, maka skor siswa secara kasar adalah relatif sama dengan hasil yang mereka peroleh pada saat pertama mereka menempuh tes tersebut (Suryadin, A & Haiyudi, 2023).

Sama seperti halnya dengan validitas, untuk mengetahui indeks reliabilitas juga digunakan rumus. Rumus yang dapat digunakan sendiri ialah *Koefisien Croanbach Alpha* (Suryadin, A & Haiyudi, 2023).

$$r_{11} = \left(\frac{n}{(n-1)} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

Keterangan:

r_{11} = reliabilitas yang dicari

$\sum \sigma_i^2$ = jumlah varians skor tiap-tiap item

σ_t^2 = varians total

N = banyaknya butir pertanyaan

Gambar 2. Koefisien Croanbach Alpha

Sementara untuk mengetahui reliabilitas sendiri dapat diamati pada bagaimana kriteria koefisiensi gambar 3 berikut.

Koefisien Korelasi	Korelasi	Interprestasi Validasi
$0,90 \leq r \leq 1,00$	sangat tinggi	sangat tepat/sangat baik
$0,70 \leq r < 0,90$	tinggi	tepat/baik
$0,40 \leq r < 0,70$	sedang	cukup tepat/cukup baik
$0,20 \leq r < 0,40$	rendah	tidak tepat/buruk
$r < 0,20$	sangat rendah	sangat tidak tepat/buruk

Gambar 3. Kriteria Reliabilitas

Reliabilitas soal STS Bahasa Indonesia kelas III SD Negeri 11

Pangkapinang dapat terlihat pada tabel 6 berikut.

Tabel 6. Reliabilitas Butir Soal Pilihan Ganda

	No soal	1	2	3	4	5	6	7	8	9	10
RELIABILITAS	Varians Soal	0,064	0,217	0,202	0,259	0,257	0,166	0,093	0,254	0,248	0,217
	$\sum \sigma_i^2$ (Varians Soal)	1,978									
	σ_i^2	2,585									
	n	10									
	n-1	9									
	r11	0,26									
	Signifikasi	0,70									
	Status	Tidak Reliabel									

Berdasarkan hasil perhitungan, reliabilitas pada tabel di atas yang menggunakan rumus VAR di *software microsoft excel*. Data diperoleh yaitu 0,26 artinya kriteria koefisiensi reabilitasnya diantara 0,20 – 0,40 yang masuk kategori rendah. Rendahnya reliabilitas soal tersebut dapat terjadi karena minimnya jumlah soal yang digunakan saat evaluasi. Selain itu, dapat juga disebabkan kondisi tes yang tidak stabil (seperti jumlah soal yang tidak memadai, tingkat kesulitan yang tidak tepat), faktor kondisi peserta tes (seperti kelelahan atau kecemasan), faktor administrasi tes (seperti kurangnya objektivitas penskoran atau instruksi yang tidak jelas), dan butir soal yang kurang representatif atau tidak jelas sehingga tidak mengukur hal yang seharusnya. Hal ini sependapat dengan Atik, F & Sukanti (2016) reliabilitas yang rendah disebabkan jumlah soal yang tidak terlalu banyak. Ada kecenderungan bahwa semakin panjang soal, dalam artian semakin banyak jumlah soal akan lebih tinggi tingkat reliabilitasnya karena semakin banyak soal maka akan semakin banyak sampel yang diukur dan proporsi jawaban yang benar semakin

banyak, sehingga faktor tebakan (*guessing*) akan semakin rendah. Selain itu, banyaknya soal yang sukar menurut siswa, karena soal yang mudah maupun sukar cenderung menghasilkan tingkat reliabilitas yang rendah. Hal ini disebabkan antara hasil tes yang mudah dan sukar keduanya berada dalam satu sebaran skor yang terbatas.

D. KESIMPULAN

Analisis butir soal Sumatif Tengah Semester (STS) materi Bahasa Indonesia kelas III di SD Negeri 11 Pangkalpinang, dapat disimpulkan melalui hasil instrumen tes butir soal dan sesuai dengan tujuan penelitian yaitu: tingkat kesukaran tidak baik, tetapi memiliki daya beda yang cukup, pengecoh soal efektif atau sudah baik, validitas soal dikategorikan tinggi dan nilai reliabilitas yang masih rendah dengan status tidak reliabel. Namun ada beberapa soal direvisi atau perlu diganti dengan butir soal yang lain terutama soal-soal yang masih kurang baik. Hasil penelitian ini diharapkan dapat menjadi rujukan guru dalam memperbaiki penyusunan soal Sumatif Tengah Semester (STS) selanjutnya.

DAFTAR PUSTAKA

- Arikunto, Suharsimi. (2018). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Arikunto, Suharsimi (2010). Penelitian Suatu Prosedur Pendekatan Praktik. Jakarta: Rineka Cipta.
- Destiniar, D., Octaria, D., & Mulbasari, A. S. (2018). Analisis Butir Soal Pilihan Ganda Dengan Aplikasi Klasika. *J-ABDIPAMAS (Jurnal Pengabdian Kepada Masyarakat)*, 2(1), 21–26.
<https://doi.org/10.30734/j-abdipamas.v2i1.180>
- Djollong, A. F. (2014). Teknik Pelaksanaan Penelitian Kuantitatif (Technique of Quantitative Research). *Istiqra'*, 2(1), 86–100.
- Farhan, M., Apriyanto, M. T., Napis, N., & Hakim, A. R. (2021). Peningkatan Kualitas Pembelajaran Melalui Pelatihan Analisis Instrumental Tes Bantuan Software Anbuso. *Jurnal Pendidikan dan pengabdian Masyarakat*, 4(4).
<https://doi.org/10.29303/jppm.v4i4.2861>
- Farida, A. M., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, 1.
- Fitriatun, Atik & Sukanti. (2016). Analisis Validitas, Reliabilitas, dan Butir Soal Latihan Ujian Ekonomi Akuntansi di MAN Maguwoharjo. *Jurnal Kajian Pendidikan Akuntansi Indonesia*. Edisi 8, 1–11.
- Janna, N. M., & Herianto. (2021). Artikel Statistik yang Benar. *Jurnal Darul Dakwah Wal- Irsyad (DDI)*, 1–12.
- Kurniawan, R. Y., Fiky Prakoso, A., Hakim, L., Mustika Dewi, R., & Widayanti, I. (2017). Pemberian Pelatihan Analisis Butir Soal Bagi Guru di Kabupaten Jombang : Efektif? *Jurnal Pemberdayaan Masyarakat Madani (JPMM)*, 1(2), 179–193.
<https://doi.org/10.21009/jpmm.001.2.03>
- Kurniawan, T. (2015). Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran IPS Sekolah Dasar. *Journal of Elementary Education*, 4(1), 1–6.
<https://journal.unnes.ac.id/sju/index.php/jee/article/view/7488>
- Magdalena, I., Anggraini, I. A., & Khoiriah, S. (2021). Analisis Daya Pembeda, dan Taraf Kesukaran pada Soal Bilangan Romawi Kelas 4 SDN Tobat 1 Balaraja. *Jurnal Pendidikan Dan Ilmu Sosial*, 3(1), 151–158.
<https://ejournal.stitpn.ac.id/index.php/nusantara>
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas Iii Sdn Karet 1 Sepatan. *BINTANG : Jurnal Pendidikan Dan Sains*, 3(2), 198–214.
<https://ejournal.stitpn.ac.id/index.php/bintang>
- Mochammad Noor Akhmadi. (2021).

- Analisis Butir Soal Evaluasi Tema 1 Kelas 4 SDN Plumbungan Menggunakan Program Anates. *Ed-Humanistics : Jurnal Ilmu Pendidikan*, 6(1), 799–806. <https://doi.org/10.33752/ed-humanistics.v6i1.1464>
- Nizary, M. A., & Kholik, A. N. (2021). Validitas Instrumen Assesmen (Analisis Validitas Isi dan Konstruk Instrumen Asesmen Buku Pelajaran Al Quran Hadis Kelas 6 Madrasah Ibtidaiyah Materi Surat Ad Dhuha Bab VI). *Jurnal Pendidikan Bahasa Arab Pengabdian Kepada Masyarakat*, 2(1), 21–42. <http://ditpsd.kemdikbud.go.id/wp-content/uploads/2017/06/Panduan-Penilaian-untuk-Sekolah-Dasar.pdf>.
- Rahmasari, D., & Ismiyati. (2016). Analisis Butir Soal Mata Pelajaran Pengantar Administrasi Perkantoran. *Economic Education Analysis Journal*, 5(1), 317–330. <http://journal.unnes.ac.id/sju/index.php/eeaj>
- Sudijono, Anas. (2017). *Pengantar Evaluasi Pendidikan*. Depok: Raja Grafindo Persada.
- Suryadin A, Maulina Hendrik, & Yorenza Meifinda. Hubungan antara Motivasi Berprestasi dan Pengetahuan Pembuatan Tes dengan Kualitas Soal Buatan Guru SMA dalam Mata Pelajaran Bahasa Indonesia dan sastra di Kota Pangkalpinang (2020). *AKSARA: Jurnal Bahasa dan Sastra* 21:2 (2020), 133 – 144. DOI: [dx.doi.org/10.23960/aksara/v21i2.pp133-144](https://doi.org/10.23960/aksara/v21i2.pp133-144) P-ISSN: 1411-2051 / E-ISSN: 2620-3928 | Universitas Lampung <http://jurnal.fkip.unila.ac.id/index.php/aksara>
- Suryadin, Asyraf., & Haiyudi. (2023). *Evaluasi Pembelajaran di Sekolah*. Purbalingga: Eureka Media Aksara
- Syamsurizal. (2020). Validitas dan Reliabilitas Alat Ukur. *Jurnal Osf*, 1–11.
- Tilaar, A. L. F., & Hasriyanti. (2019). Analisis Butir Soal Semester Ganjil Mata Pelajaran Matematika pada Sekolah Menengah Pertama. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 8(1), 57–68. <https://doi.org/10.15408/jp3i.v8i1.13068>