

PROFIL VALIDITAS DAN RELIABILITAS BUTIR SOAL MATEMATIKA UJIAN AKHIR SEMESTER KELAS VIII SMP DI BANDA ACEH

Intan Kemala Sari¹⁾, Nurul Fajri²⁾, Sri Mulyani³⁾

^{1) 2) 3)} STKIP Bina Bangsa Getsmpena

e-mail: intankemalasari00@gmail.com

Abstrak

Evaluasi pembelajaran merupakan hal yang mutlak dilakukan setelah proses pembelajaran telah diselesaikan. Hal ini dilakukan untuk mengukur keberhasilan proses belajar mengajar, merencanakan tindak lanjut, dan melakukan perbaikan. Namun apa jadinya jika instrument evaluasi tidak dapat mengukur keberhasilan proses yang telah dijalankan? Penelitian ini bertujuan untuk mengetahui bagaimana profil validitas dan reliabilitas butir soal tes ujian semester pelajaran matematika yang digunakan guru untuk mengukur keberhasilan proses belajar mengajarnya. Penelitian ini dilakukan dengan cara menganalisis butir soal ujian matematika siswa yang terdiri dari 40 soal pilihan ganda dan 5 soal uraian. Hasil analisis menunjukkan bahwa 40% soal pilihan ganda valid namun tidak satu pun soal yang reliable. Tingkat validitas pun termasuk dalam kategori sedang dan rendah. Sedangkan satu dari lima soal uraian tidak valid namun selebihnya memenuhi syarat valid dan reliable dengan kategori sangat tinggi. Diketahui guru sering mengulang pemberian soal yang sama dari tahun ke tahun. Namun dengan adanya uji validitas dan reliabilitas ini diharapkan guru dapat melakukan revisi yang relevan terhadap soal yang tidak valid dan reliable, serta memasukkan soal yang valid dan reliabel menjadi bank soal pribadi guru dan dapat menjadi masukan jika dibutuhkan soal tes yang memadai.

Kata Kunci: *validitas, reliabilitas, butir soal*

Abstract

Evaluation of learning is activity that is done after the learning process is completed. This activity have to do to measure the success of the teaching and learning process in the class, follow-up, and make improvements to the next learning activity. But what happens if the evaluation instrument cannot measure the success of the process that has been carried out? This study aims to study how the profile of validity and reliability of semester exam questions in mathematics is used by the teachers to measure the learning process of teaching and learning. This research was conducted by analyzing mathematical questions consisting of 40 multiple choice questions and 5 description questions. The results of the analysis show that 40% of multiple choice questions are valid but not same condition with reliability score. The level of validity is included in the medium and low category. While one of the five description questions is invalid but the rest fulfill valid and reliable requirements with very high categories. It is known that teachers often repeat the same questions year after year. However, by asking questions about validity and reliability, it is expected that the teacher can make relevant revisions to invalid and reliable questions, and ask valid and reliable questions to become the teacher's personal question bank and can be a necessary problem.

Keywords: *validity, reliability, exam questions*

PENDAHULUAN

Matematika merupakan salah satu bidang ilmu yang memiliki banyak penerapan dalam kehidupan sehari-hari. Pendidikan matematika dapat diartikan sebagai proses perubahan baik kognitif, afektif, maupun psikomotorik ke arah kematangan sikap seseorang sesuai dengan kebenaran logika (Kurikulum 2013). Matematika juga merupakan ilmu pengetahuan yang memiliki pola keteraturan yang berhubungan dengan ide, proses, dan penalaran.

Proses belajar mengajar memiliki dampak yang besar terhadap kehidupan seseorang. Dalam kehidupan bermasyarakat, manusia dituntut berpikir kritis, logis dan kreatif. Hal ini diperlukan untuk memudahkan manusia untuk berinteraksi dalam kehidupan sosial. Oleh karena itu, matematika tidak bisa dipisahkan dari kehidupan kita. Agar tercapainya hasil belajar sesuai dengan yang diinginkan dan menjadi lebih baik, maka perlu memperbaiki proses belajar mengajar di sekolah. Upaya peningkatan kualitas proses belajar mengajar dan tercapainya hasil belajar sesuai dengan harapan merupakan salah satu bagian dari peningkatan kualitas pendidikan.

Menurut Arikunto (2011) guru maupun pendidik lainnya perlu mengadakan penilaian terhadap hasil belajar siswa karena dalam dunia pendidikan, khususnya dunia persekolahan penilaian hasil belajar mempunyai makna yang penting, baik bagi siswa, guru maupun sekolah. Setiap teknik penilaian memerlukan instrumen penilaian. Instrumen penilaian merupakan alat bantu yang digunakan oleh guru atau penilai untuk mengumpulkan data tentang karakteristik siswa dengan cara melakukan pengukuran.

Dengan melakukan pengukuran akan diperoleh data yang objektif yang diperlukan untuk menilai hasil belajar siswa. Selain diperoleh data yang objektif, dengan menggunakan instrumen maka pekerjaan penilaian menjadi lebih mudah dan hasilnya lebih baik, lebih cermat, lengkap dan sistematis. Untuk memperoleh data pengukuran hasil belajar siswa yang baik diperlukan instrumen yang berkualitas, yaitu yang memenuhi syarat validitas dan reliabilitas.

Evaluasi merupakan suatu proses sistematis dalam memeriksa, menentukan, membuat keputusan atau menyediakan informasi terhadap program yang telah dilakukan dan sejauh mana sebuah program tercapai. Tanpa evaluasi, maka tidak akan diketahui bagaimana kondisi objek evaluasi tersebut dalam rancangan, pelaksanaan serta hasilnya. Menurut Yunanda (2009) pengertian istilah "evaluasi merupakan kegiatan yang terencana untuk mengetahui keadaan sesuatu objek dengan menggunakan instrumen dan hasilnya dibandingkan dengan tolak ukur untuk memperoleh kesimpulan".

Untuk memastikan pelaksanaan suatu program mencapai sasaran dan tujuan yang direncanakan, maka perlu diadakan evaluasi dalam rangka peningkatan kinerja program tersebut. Menurut Sukardi (2014: 1) mengemukakan pandangannya bahwa evaluasi adalah proses mencari data atau informasi tentang objek atau subjek yang dilaksanakan untuk tujuan pengambilan keputusan terhadap subjek ataupun objek tersebut.

Menurut Arifin (2011: 117) Banyak alat atau instrumen yang dapat digunakan dalam kehidupan evaluasi. Salah satunya adalah tes. Istilah tes tidak hanya populer di lingkungan persekolahan, tetapi juga di

luar sekolah bahkan di masyarakat umum. Kita sering mendengar istilah tes kesehatan, tes olahraga, tes makanan, tes kendaraan dan lain-lain. Di sekolah juga sering kita dengar istilah pretes, postes, tes formatif, tes sumatif, dan sebagainya. Di sekolah, tes ini sering juga disebut dengan tes prestasi belajar, tes ini banyak digunakan untuk mengukur prestasi belajar peserta didik dalam bidang kognitif, seperti pengetahuan, pemahaman, aplikasi, analisis, sintesis, dan evaluasi. Penggunaan tes dalam dunia pendidikan sudah dikenal sejak dahulu kala, sejak orang mengenal pendidikan itu sendiri. Artinya, tes mempunyai makna tersendiri dalam dunia pendidikan, khususnya dalam pembelajaran.

Proses untuk menuju suatu tujuan atau suatu keberhasilan yang telah ditetapkan sangat dipengaruhi oleh beberapa faktor. Menurut Arifin (2011: 1), antara lain faktor guru terhadap peserta didik dalam proses belajar mengajar dikelas. Maka setiap guru dituntut agar mempunyai tanggung jawab dalam merencanakan dan melaksanakan evaluasi. Salah satu alat evaluasi yang digunakan adalah tes. Dalam mencapai tujuan pembelajaran, diperlukan soal tes yang memiliki kualitas. Soal tes yang berkualitas adalah soal tes yang dapat berfungsi dengan baik dan efektif dalam mengukur kemampuan peserta didik. Analisis butir soal atau analisis item merupakan pengkajian pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai (Sudjana, 2011:135). Analisis butir soal yang dilakukan akan dapat meningkatkan kualitas soal melalui unsur validitas dan reliabilitas, semakin tinggi nilai validitas dan reliabilitas suatu instrumen, maka

akan semakin tepat data yang didapat dari suatu penelitian.

Menurut Arifin (2011:30) terdapat beberapa prinsip umum yang perlu diperhatikan dalam melakukan evaluasi, yaitu:

a. Kontinuitas

Evaluasi tidak boleh dilakukan secara insidental karena pembelajaran itu sendiri adalah suatu proses yang kontinu. Oleh sebab itu, evaluasi pun harus dilakukan secara kontinu. Hasil evaluasi yang diperoleh pada suatu waktu harus senantiasa dihubungkan dengan hasil-hasil pada waktu sebelumnya, sehingga dapat diperoleh gambaran yang jelas dan berarti tentang perkembangan peserta didik. Perkembangan belajar peserta didik tidak dapat dilihat dari dimensi produk saja, tetapi juga dimensi proses bahkan dari dimensi input.

b. Komprehensif

Dalam melakukan evaluasi terhadap suatu objek, guru harus mengambil seluruh objek itu sebagai bahan evaluasi. Misalnya, jika objek evaluasi itu adalah peserta didik, maka seluruh aspek kepribadian peserta didik itu harus dievaluasi, baik yang menyangkut kognitif, afektif maupun psikomotor. Begitu juga dengan objek-objek evaluasi yang lain.

c. Adil dan Objektif

Dalam melaksanakan evaluasi, guru harus berlaku adil tanpa pilih kasih. Kata "adil" dan "objektif" memang mudah diucapkan, tetapi sulit dilaksanakan. Meskipun demikian, kewajiban manusia adalah harus berikhtiar. Semua peserta didik harus diberlakukan sama tanpa "pandang bulu". Guru juga hendaknya bertindak secara objektif, apa adanya sesuai

dengan kemampuan peserta didik. Oleh sebab itu, sikap *like and dislike*, perasaan, keinginan, dan prasangka yang bersifat negatif harus dijauhkan. Evaluasi harus didasarkan atas kenyataan (data dan fakta) yang sebenarnya, bukan hasil manipulasi atau rekayasa.

d. Kooperatif

Dalam kegiatan evaluasi guru hendaknya bekerja sama dengan semua pihak, seperti orang tua peserta didik, sesama guru, kepala sekolah, termasuk peserta didik itu sendiri. Hal ini dimaksudkan agar semua pihak merasa puas dengan hasil evaluasi, dan pihak-pihak tersebut merasa dihargai.

e. Praktis

Praktis mengandung arti mudah digunakan, baik oleh guru itu sendiri yang menyusun alat evaluasi maupun orang lain yang menggunakan alat tersebut. Untuk itu harus diperhatikan bahasa dan petunjuk mengerjakan soal.

Menurut Sudijono (2012: 59) merinci kegiatan evaluasi hasil belajar kedalam enam langkah pokok yaitu:

a. Menyusun rencana evaluasi hasil belajar
Sebelum evaluasi hasil belajar dilaksanakan, harus disusun terlebih dahulu perencanaannya secara baik dan matang. Perencanaan evaluasi hasil belajar itu umumnya mencakup enam jenis kegiatan, yaitu:

- ✓ Merumuskan tujuan dilaksanakannya evaluasi
- ✓ Menetapkan aspek-aspek yang akan dievaluasi, misalnya aspek kognitif, aspek afektif, atau aspek psikomotorik
- ✓ Memilih dan menentukan teknik yang akan digunakan di dalam pelaksanaan evaluasi, misalnya

apakah evaluasi itu akan dilaksanakan dengan teknik tes atau non tes

- ✓ Menyusun alat-alat pengukur yang akan dipergunakan dalam pengukuran dan penilaian hasil belajar
- ✓ Menentukan tolak ukur, norma atau kriteria yang akan dijadikan pegangan atau patokan dalam memberikan interpretasi terhadap hasil evaluasi
- ✓ Menentukan frekuensi dari kegiatan evaluasi hasil belajar itu sendiri.

b. Menghimpun data

Langkah selanjutnya adalah menghimpun data dalam evaluasi belajar yang berarti melaksanakan pengukuran. Pengukuran data disesuaikan dengan jenis data yang ingin diambil misalnya dengan menyelenggarakan tes hasil belajar (apabila evaluasi hasil belajar itu menggunakan teknik tes) atau melakukan pengamatan, wawancara atau angket dengan menggunakan instrument-instrumen tertentu berupa rating scale, check list, interview guide atau questionnaire (apabila evaluasi hasil belajar itu menggunakan teknik non tes).

c. Melakukan verifikasi data

Data yang telah berhasil dihimpun harus disaring terlebih dahulu sebelum diolah lebih lanjut. Proses penyaringan itu dikenal dengan istilah penelitian data atau verifikasi data. Verifikasi data yang dimaksud untuk dapat memisahkan data yang baik dari yang kurang baik. Data yang baik yaitu data yang dapat memperjelas gambaran yang akan diperoleh mengenai diri individu atau sekelompok individu yang sedang

dievaluasi. Sedangkan yang dimaksud data yang kurang baik yaitu data yang akan mengaburkan gambaran yang akan diperoleh apabila data itu ikut serta diolah.

d. Mengolah dan menganalisis data

Mengolah data dan menganalisis hasil evaluasi dilakukan dengan maksud untuk memberikan makna terhadap data yang telah berhasil dihimpun dalam kegiatan evaluasi. Dalam mengolah dan menganalisis data itu dapat dilakukan dengan menggunakan teknik statistik.

e. Memberi interpretasi dan menarik kesimpulan

Penafsiran atau interpretasi terhadap data hasil evaluasi belajar pada hakikatnya adalah merupakan hasil verbalisasi dari makna yang terkandung dalam data yang telah mengalami pengolahan dan penganalisisan tersebut. Atas dasar interpretasi terhadap data evaluasi hasil belajar itu pada akhirnya dapat dikemukakan kesimpulan.

f. Tindak lanjut hasil evaluasi

Bertitik tolak dari data hasil evaluasi yang telah disusun, diatur, diolah, dianalisis, dan disimpulkan sehingga dapat diketahui apa makna yang terkandung didalamnya maka pada akhirnya evaluator akan dapat mengambil keputusan atau merumuskan kebijakan-kebijakan yang dipandang perlu sebagai tindak lanjut dari kegiatan evaluasi tersebut.

Menurut Arikunto (2013: 72) sebuah tes dapat dikatakan baik sebagai alat pengukur, harus memenuhi persyaratan tes, yaitu:

a. Validitas

Validitas sebuah tes disebut valid apabila tes tersebut dapat dengan tepat mengukur apa yang hendak diukur. Misalnya, untuk mengukur prestasi siswa dalam proses belajar mengajar, bukan diukur melalui nilai yang diperoleh pada waktu ulangan akan tetapi pada ketepatan menjawab pertanyaan yang diajukan guru yang berarti relevan pada permasalahannya.

b. Reliabilitas

Reliabilitas adalah dapat dipercaya, suatu tes dikatakan dapat dipercaya jika tes tersebut memberikan hasil yang tetap apabila diteskan berkali-kali pada peserta tes yang sama dalam waktu yang berbedah.

c. Objektivitas

Suatu tes dikatakan memiliki objektivitas apabila dalam melaksanakan tes itu tidak ada faktor subjektif atau unsur pribadi yang mempengaruhi. Apabila dikaitkan dengan reliabilitas maka objektivitas menekankan ketetapan pada sistem penilaiannya, sedangkan reliabilitas menekankan ketetapan dalam hasil tes.

d. Praktikabilitas

Sebuah tes dikatakan memiliki praktikabilitas yang tinggi apabila tes tersebut bersifat praktis dan yang mudah dilaksanakan, mudah pemeriksaannya, dan dilengkapi dengan petunjuk-petunjuk yang jelas.

e. Ekonomis

Ekonomis yang dimaksud disini adalah bahwa pelaksanaan tes tersebut tidak membutuhkan ongkos atau biaya yang mahal, tenaga yang banyak, dan waktu yang lama.

Berdasarkan uraian di atas, maka penulis tertarik untuk melakukan

penelitian untuk menganalisis Validitas dan Reliabilitas Butir Soal Matematika Ujian Akhir Semester Kelas VIII SMP.

METODE PENELITIAN

Menurut Surapranata (2009: 10) Analisis soal secara kuantitatif menekankan pada analisis karakteristik internal tes melalui data yang diperoleh secara empiris. Adapun tujuan dari analisis butir soal adalah untuk meningkatkan kualitas butir soal yang akan digunakan dalam tes. Penelitian ini dilakukan dengan cara menganalisis soal dan jawaban siswa menggunakan pendekatan kuantitatif dimana penjabarannya menggunakan metode deskriptif karena data yang diperoleh dalam bentuk angka-angka.

Teknik yang digunakan untuk mengumpulkan data dalam penelitian ini adalah dokumentasi, dokumentasi yang dimaksud adalah berupa tes soal ujian kenaikan kelas mata pelajaran matematika yaitu tes tulis yang berupa butir soal pilihan ganda dan essay, Butir soal atau tes merupakan alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dalam suasana, dengan cara dan aturan-aturan yang sudah ditentukan. Tes hasil belajar adalah sekelompok pertanyaan atau tugas-tugas yang harus dijawab atau diselesaikan oleh siswa dengan tujuan untuk mengukur kemajuan belajar siswa.

Peneliti mengambil tes soal yang sudah di jawab oleh siswa dari sekolah, setelah tes soal sudah terkumpul langkah selanjutnya peneliti menganalisis soal

tersebut dengan menentukan valid dan reliabelnya soal yang telah dianalisis, tes soal yang dianalisis tersebut berupa tes soal cois yang jumlahnya 40 butir dan essay yang jumlahnya 5 butir soal, jadi jumlah soalnya adalah 45 butir soal.

Teknik analisis data yang digunakan dalam penelitian ini adalah metode analisis yaitu menganalisis dari soal tes Ujaian Akhir Semester dan lembar jawaban yang telah di isi oleh siswa.

1. Validitas

Validitas tes merupakan tingkat keabsahan atau ketepatan suatu tes yang benar-benar mengukur apa yang hendak diukur, dengan kata lain validitas tes dilakukan untuk menunjukkan tingkat ketepatan suatu instrument tes dalam mengukur sasaran yang hendak diukur (Arikunto, 2013: 80). Uji validitas tes yang digunakan adalah teknik korelasi *product moment* yang dikemukakan oleh pearson (dalam Arkunto, 2013: 87).

$$r_{xy} = \frac{n \sum xy - (\sum x) (\sum y)}{\sqrt{(n \sum x^2 - (\sum x)^2) (n \sum y^2 - (\sum y)^2)}}$$

Jika r -hitung $\geq r$ -tabel, maka instrument atau item-item pertanyaan berkorelasi signifikan terhadap skor total (dinyatakan valid). Jika r -hitung $< r$ -tabel, maka instrument tidak berkorelasi signifikan terhadap skor total (dinyatakan tidak valid). Interpretasi untuk besarnya koefisien korelasi dapat dilihat pada tabel berikut

Tabel 1. Kategori Validitas Soal

Batasan	Kategori
$0,80 < r_{xy} \leq 1,00$	Sangat Tinggi
$0,60 < r_{xy} \leq 0,80$	Tinggi
$0,40 < r_{xy} \leq 0,60$	Cukup
$0,20 < r_{xy} \leq 0,40$	Rendah
$0,00 < r_{xy} \leq 0,20$	Sangat Rendah

(Sumber: Arikunto, 2013: 92)

2. Reliabilitas

Reliabilitas adalah kestabilan skor yang diperoleh ketika diuji ulang dengan tes yang sama pada situasi yang berbeda atau satu pengukuran kepengukuran lainnya. Untuk menentukan reliabilitas tes, di gunakan rumus varians butir.

$$\alpha^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}$$

Pemberian interpretasi terhadap koefisien reliabilitas tes pada umumnya digunakan patokan sebagai berikut: Apabila nilai α sama dengan atau lebih besar daripada 0,70 berarti tes hasil belajar yang sedang diuji reliabilitasnya dinyatakan telah memiliki reliabilitas yang tinggi. Apabila nilai α lebih kecil dari pada 0,70 berarti bahwa tes tes hasil belajar yang sedang diuji reliabilitasnya dinyatakan belum memiliki reliabilitas yang tinggi.

HASIL DAN PEMBAHASAN

Dalam data yang diperoleh pada penelitian ini adalah lembaran jawaban tes ujian akhir semester mata pelajaran matematika kelas VIII semester II (genap) yang terdiri dari 18 siswa. Soal tes tersebut berjumlah 45 soal yang terdiri 40 soal pilihan ganda dan 5 soal uraian. Data nilai ujian tersebut diperoleh dari guru pada SMP Negeri 9 Banda Aceh tahun 2017/2018.

Lembar jawaban siswa diperoleh dari guru bidang studi matematika kelas VIII SMP Negeri 9 Banda Aceh setelah siswa selesai mengikuti ujian akhir semester.

Lembar jawaban siswa tersebut telah didokumentasikan dalam raport kenaikan kelas. Berdasarkan lembar jawaban siswa tersebut, selanjutnya ditabulasikan untuk setiap pilihan jawaban yang dijawab siswa.

Sebelum menganalisis butir soal, maka terlebih dahulu ditentukan skor untuk masing-masing butir soal. Untuk soal pilihan ganda (soal no 1-40), setiap jawaban yang benar akan diberi skor 1, dan jawaban yang salah diberi skor 0.

Untuk soal uraian pada nomor 1 akan diberi skor 14 untuk jawaban yang benar, pada soal nomor 2 akan diberi skor 10 untuk jawaban yang benar, pada soal nomor 3 akan diberi skor 6 untuk jawaban yang benar, pada soal nomor 4 akan diberi skor 20 untuk jawaban yang benar, dan pada soal nomor 5 akan diberi skor 10 untuk jawaban yang benar.

Selanjutnya dibuat tabulasi jawaban siswa setelah pemberian skor kemudian baru ditentukan validitas dan reliabilitas masing-masing soal tersebut. Berikut hasil perhitungan validitas dan reliabilitas soal ujian akhir semester genap mata pelajaran matematika SMP.

VALIDITAS PILIHAN GANDA

■ Tidak valid ■ Valid



Gambar 1. Validitas Soal Pilihan Ganda

Hasil analisis secara keseluruhan menunjukkan 12 butir soal atau 40% dari soal secara keseluruhan termasuk dalam kategori soal yang valid dan 28 butir soal atau 60% dari soal keseluruhan tidak valid. Ditinjau dari kevalidan dalam kategori valid $\alpha = 0,05$ dan $r_{tabel} = 0,468$ diketahui bahwa validitas butir-butir soal pilihan ganda pada ujian akhir semester genap kelas VIII SMP di Banda Aceh Tahun pelajaran 2017/2018 bidang studi matematika tergolong rendah dan sedang.

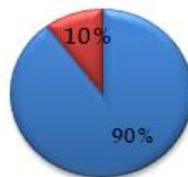
Sedangkan pada analisis reliabilitas butir soal tes ujian akhir semester genap tahun 2017/2018 bidang studi matematika

kelas VIII SMP secara keseluruhan menunjukkan tidak reliabel, karena ke 40 butir soal tersebut tidak ada satupun yang reliabel, soal pilihan ganda dikategorikan reliabelnya rendah dan sangat.

Hasil analisis validitas soal uraian kevalidan dalam kategori valid $\alpha = 0,05$ dan $r_{tabel} = 0,468$ terlihat bahwa validitas butir-butir soal uraian pada ujian akhir semester genap kelas VIII SMP Tahun pelajaran 2017/2018 bidang studi matematika tergolong sangat tinggi. Adapun proporsi butir soal yang valid dan tidak valid dapat dilihat pada diagram berikut.

Validitas Soal Uraian

■ Valid ■ Tidak valid



Gambar 2. Validitas Soal Uraian

Hasil analisis keseluruhan menunjukkan 4 butir soal atau 90% dari soal secara keseluruhan termasuk dalam kategori soal yang valid dan 1 butir soal atau 10% dari soal keseluruhan masih belum valid.

Sedangkan hasil analisis reliabilitas soal uraian secara keseluruhan

menunjukkan reliabel, karena ke 5 butir soal tersebut tidak ada satupun yang tidak reliabel, soal uraian dikategorikan reliabelnya sangat tinggi.

Berdasarkan hasil analisis validitas dan reliabilitas butir soal baik pilihan ganda maupun uraian dapat dikatakan bahwa soal yang diberikan untuk

mengevaluasi proses belajar mengajar dalam satu semester cukup memadai namun perlu ditingkatkan. Misalnya dalam soal pilihan ganda, tidak terdapat satu soal pun yang reliabel. Padahal nilai reliabilitas suatu butir soal menunjukkan tingkat konsistensi dari suatu soal sehingga dapat dipercaya kebenarannya. Pengukuran memiliki reliabilitas yang tinggi jika pengukuran tersebut mampu menghasilkan data yang reliabel. Suatu tes dikatakan reliabel jika tes tersebut selalu memberikan hasil yang sama bila diberikan pada kelompok yang sama dalam waktu dan kesempatan yang berbeda. Dengan ketidakreliabel butir soal yang diberikan, berarti menunjukkan bahwa sebaiknya soal tidak diberikan lagi pada kesempatan evaluasi yang lain. Jika ingin mengulang penggunaan butir soal yang sama tersebut, ada baiknya untuk direvisi terlebih dahulu.

Sedangkan validitas dalam soal baik pilihan ganda maupun uraian sudah masuk dalam kategori yang memadai yaitu diatas 40%. Hal ini menunjukkan bahwa suatu ukuran butir soal yang menunjukkan tingkat kesahihan suatu tes. Suatu tes dikatakan valid apabila tes tersebut mengukur apa yang hendak diukur. Tes memiliki validitas yang tinggi jika hasilnya sesuai dengan kriteria, dalam arti memiliki kesejajaran antara tes dan kriteria (Arikunto, 1999: 65). Dalam analisis butir soal di penelitian ini, analisis validitasnya masuk dalam kategori sedang. Dengan demikian, jika hanya mempertimbangkan pengukuran apa yang hendak diukur, maka butir soal yang diberikan sudah cukup memadai.

Dalam hampir dua dekade terakhir kemampuan matematika siswa Indonesia mengalami posisi stagnan yang relatif rendah. Dalam studi yang serupa di Banda

Aceh, diketahui bahwa terdapat indikasi kemunduran perkembangan kognitif siswa SMP yang seharusnya sudah memasuki tahap abstraksi awal namun masih pada memandang suatu masalah berdasarkan satu sudut pandang (Sari, 2018). Selain pengaruh proses belajar mengajar yang kurang memberikan kesempatan kepada siswa untuk mengkonstruksi pemahaman matematika formal dari matematika informal berbasis masalah, instrument evaluasi yang disajikan pun dianggap cukup berat atau dengan kategori yang cukup tinggi.

Di sisi lain, Indonesia harus mengejar ketertinggalannya dalam ajang pengukuran prestasi belajar siswa berdasarkan standar internasional seperti TIMSS atau PISA, namun dilema bagi tenaga pendidikan adalah menyesuaikan antara konsep yang dibangun dengan instrument evaluasi yang harus dihadapi siswa. Oleh sebab itu pemerintah selalu mengupayakan tes akhir kelulusan siswa secara nasional yang harus diseimbangkan dengan standar internasional namun harus mempertimbangkan keadaan siswa di Indonesia dari berbagai sosio-kultural dan letak geografis yang sedikit banyak akan mempengaruhi hasil yang ditetapkan.

Oleh sebab itu perlu kerja keras dan kerjasama yang baik dengan intensitas tinggi antara susunan proses belajar mengajar, standar ukuran yang relevan dan relatif tinggi, serta instrument evaluasi yang valid, praktis, reliabel, dan dapat dipahami dengan baik.

SIMPULAN DAN SARAN

Penelitian yang telah dilakukan pada SMP di Banda Aceh kelas VIII semester genap dengan menerapkan analisis tiap butir soal pilihan ganda menghasilkan 40% kategori valid dan 60%

kategori tidak valid, sedangkan butir soal uraian menghasilkan 90% kategori valid dan 10% kategori tidak valid. Memiliki reliabilitas yang rendah untuk butir soal pilihan ganda, dan memiliki reliabilitas yang sangat tinggi untuk butir soal uraian.

Soal pilihan ganda perlu adanya revisi karena tidak ada satupun yang bisa digunakan karena valid tapi tidak adav soal reliabel, jadi soal tersebut perlu adanya revisi kembali. Jadi soal pilihan ganda tersebut kevalidannya rendah atau tidak valid, mungkin dikarenakan soal 40 butir itu terlalu banyak untuk waktu 2 jam atau 80 menit pada jam pelajaran Sekolah Menengah Pertama sehingga peserta didik tidak cukup waktu untuk menjawab soal dan ada juga soal yang tidak jelas pernyataanya, soal yang tidak ada jawaban sehingga soal tersebut tidak terjawab.

Sebanyak 5 butir soal uraian dari keseluruhan butir soal dikategorikan valid ada 4 soal dengan persentase 90%, butir

soal yang valid nomor 1, 2, 4, 5. Sedangkan butir soal yang tidak valid ada 1 soal dengan persentase 10%, butir soal yang tidak valid nomor 3. Jadi butir soal uraian ada 4 butir soal yang dapat disimpan untuk digunakan kembali. Soal ujian pilihan ganda termasuk soal yang memiliki reliabilitas yang rendah sedangkan soal ujiian uraian termasuk soal yang memiliki reliabilitas yang sangat tinggi.

Kepada guru bidang studi matematika disarankan untuk melakukan uji coba terlebih dahulu pada setiap butir soal yang akan digunakan sebagai soal ujian akhir semester. Sehingga akan didapat soal ujian yang mampu menjadi tolak ukur penilaian atas hasil belajar siswa sekaligus untuk meningkatkan pencapaian tujuan pembelajaran itu sendiri. Pada soal yang tergolong baik dapat dijadikan bank soal untuk nantinya digunakan kembali dan soal yang kurang baik sebaiknya dilakukan revisi.

DAFTAR PUSTAKA

- Arifin, Z. 2011. *Evaluasi Pembelajaran, Prinsip Teknik Pembelajaran*. Bandung: Remaja Rosdakarya.
- Arikunto, S. 2013b. *Prosedur Penelitian Suatu Pendekatan Praktik*. Rineka Cipta, Jakarta.
- Arikunto, S. 2009. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, S. 2011a. *Dasar-Dasar Evaluasi Pendidikan*. Edisi Revisi, Cetakan kesebelas, Jakarta: Bumi Aksara.
- Sari, I. K., Nasriadi, A., & Salmina, M. (2018, September). Students' understanding of charts: the study of PISA's problem-solving in the content of data. In *Journal of Physics: Conference Series* (Vol. 1088, No. 1, p. 012108). IOP Publishing.
- Sukardi. 2014. *Metodologi Penelitian Pendidikan Kompetensi dan Praktikan*. Jakarta: Bumi Aksara.
- Surapranata, S. 2009. *Analisis, Validitas, Reliabilitas dan Interpretasi Tes*, Bandung: PT. Remaja Rosda Karya.
- Yunanda, Martha. 2009. *Evaluasi Dalam Islam*. [Http//id.shvoong.com/social-sciences/education/1956775-evaluasi-dalam-islam/](http://id.shvoong.com/social-sciences/education/1956775-evaluasi-dalam-islam/). Diakses pada 01 maret 2010