

BAB V

KAJIAN DAN SARAN

A. Kajian Produk yang Telah Direvisi

1. Proses Pengembangan Soal HOTS Matematika Berkonteks Budaya *Nyadran* Desa Sonoageng

Penelitian dan pengembangan yang dilakukan peneliti telah menghasilkan produk berupa Soal *Higher Order Thinking Skills* (HOTS) Matematika Berkonteks Budaya *Nyadran* Desa Sonoageng Kabupaten Nganjuk yang berjumlah 20 butir soal dengan bentuk soalnya pilihan ganda. Soal yang dikembangkan bertujuan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik yang meliputi kemampuan menganalisis, mengevaluasi, dan mencipta.

Tahap penyusunan spesifikasi tes, pada tahap ini peneliti melakukan analisis secara mendalam terhadap unsur-unsur matematis yang terkandung dalam budaya *Nyadran* di Desa Sonoageng, dimulai dari menganalisis keterkaitan unsur budaya *Nyadran* dengan aktivitas matematis kemudian peneliti melakukan analisis persetujuan rater untuk menilai hasil analisis keterkaitan unsur budaya dengan aktivitas matematis yang telah dikaji oleh peneliti. Tujuan dari analisis persetujuan rater ini adalah produk soal yang disusun tetap mencerminkan realitas budaya yang sebenarnya dan tidak sekadar menjadi tempelan konteks. Berdasarkan hasil analisis unsur matematis

dalam budaya *Nyadran* langkah selanjutnya adalah menyusun kisi-kisi instrumen tes.

Tahap penulisan butir soal, pada tahap ini peneliti membuat butir soal dengan mengacu pada kisi-kisi yang telah disusun pada tahap sebelumnya. Dalam proses penulisan butir soal, terdapat beberapa hal penting yang menjadi perhatian yaitu, kesesuaian dengan indikator soal, kesesuaian dengan level kognitif khususnya pada level HOTS (C4, C5, dan C6), kejelasan dan ketepatan penggunaan bahasa, kaidah penulisan soal pilihan ganda yang baik dan benar, dan keautentikan konteks budaya *Nyadran*.

Tahap penelaahan butir soal, pada tahap ini peneliti melakukan penelaahan butir soal dengan validasi ahli (*expert judgement*) oleh dua dosen Tadris Matematika. Proses validasi ini dilakukan berdasarkan tiga aspek penilaian, yaitu materi, konstruksi, dan bahasa, yang masing-masing memiliki sejumlah indikator untuk menilai kesesuaian, kejelasan, serta kualitas butir soal. Hasil dari validasi tersebut menunjukkan bahwa soal HOTS yang dikembangkan sudah memenuhi kriteria sangat valid dan valid, sehingga butir soal yang telah divalidasi bisa dilanjutkan pada tahap uji coba dengan sedikit revisi.

Tahap pelaksanaan uji coba, tahap uji coba dilakukan untuk menguji kualitas soal HOTS matematika yang telah dikembangkan dengan melibatkan 25 siswa kelas XI-1 MAN 3 Nganjuk sebagai responden. Melalui uji coba ini, peneliti memperoleh data empiris yang digunakan

untuk menilai karakteristik butir soal secara objektif. Hasil dari tahap ini menjadi dasar penting dalam menentukan kelayakan setiap butir soal, apakah perlu dipertahankan, direvisi, atau dieliminasi sebelum digunakan pada skala yang lebih luas.

Tahap analisis butir soal, pada tahap ini peneliti menganalisis reliabilitas butir soal HOTS matematika menggunakan dua pendekatan, yaitu *Model Rasch* dengan bantuan perangkat lunak *Ministeps* dan Teori Tes Klasik. *Model Rasch* memberikan informasi rinci terkait reliabilitas, kesesuaian butir soal, tingkat kesukaran, serta abilitas individu, sedangkan Teori Tes Klasik memberikan gambaran umum melalui perhitungan reliabilitas tes dan tingkat kesukaran. Penggunaan kedua pendekatan ini bertujuan untuk memperoleh analisis yang lebih komprehensif karena masing-masing memiliki keunggulan dalam menilai kualitas instrumen dari sudut pandang yang berbeda.

Tahap revisi butir soal, pada tahap ini peneliti melakukan revisi butir soal melalui dua tahap, yaitu berdasarkan validasi ahli dan hasil analisis empiris. Revisi awal bertujuan memastikan kesesuaian soal dari aspek materi, konstruksi, dan bahasa agar memenuhi kriteria sangat valid sebelum diujicobakan. Selanjutnya, revisi dilakukan berdasarkan hasil uji coba menggunakan *Model Rasch* dan tes klasik, yang mencakup perbaikan terhadap reliabilitas, kesesuaian butir soal, tingkat kesukaran, dan abilitas peserta didik. Perbaikan dilakukan dengan merevisi redaksi, konteks, tingkat kesukaran, hingga eliminasi butir yang tidak layak.

2. Validitas Butir Soal

Validitas merupakan suatu proses pembuktian yang berkelanjutan yang berkaitan dengan sejauh mana bukti empiris dan landasan teoretis mampu mendukung ketepatan interpretasi terhadap skor tes sesuai dengan tujuan pengukuran. Proses validasi dilakukan melalui pengumpulan berbagai bukti yang berfungsi sebagai dasar ilmiah dalam menafsirkan hasil tes (Sumintono & Widhiarso, 2015). Dengan demikian, validitas tidak terletak pada instrumen tes itu sendiri, melainkan pada interpretasi terhadap skor yang dihasilkan, karena validitas tidak bergantung pada bentuk atau jenis tes, melainkan pada ketepatan penafsiran skor tersebut. Oleh sebab itu, apabila suatu skor tes digunakan atau ditafsirkan dalam lebih dari satu cara, maka setiap bentuk interpretasi tersebut harus melalui proses validasi agar dapat dipertanggungjawabkan secara ilmiah (Sumintono & Widhiarso, 2015).

Keberadaan validitas dalam suatu penelitian sangat penting karena menjadi dasar dalam menjamin keakuratan dan kepercayaan terhadap hasil penelitian. Tanpa adanya validitas, hasil pengukuran dapat menimbulkan interpretasi yang keliru, sehingga kesimpulan penelitian menjadi tidak tepat (Sumintono & Widhiarso, 2015). Dalam konteks pengembangan soal HOTS, validitas memastikan bahwa butir soal benar-benar mengukur kemampuan berpikir tingkat tinggi, bukan sekadar kemampuan mengingat atau memahami. Sebaliknya, apabila suatu instrumen tidak melalui proses validasi, maka terdapat risiko besar bahwa data yang dihasilkan tidak mencerminkan kondisi sebenarnya.

Hal ini dapat menyebabkan bias dalam pengukuran, kesalahan dalam pengambilan keputusan, serta menurunnya kualitas penelitian secara keseluruhan. Bahkan, tanpa validitas, hasil penelitian menjadi diragukan keabsahannya dan sulit untuk dijadikan dasar dalam pengembangan ilmu pengetahuan maupun praktik pendidikan (Sumintono & Widhiarso, 2015).

Proses validasi terhadap butir soal HOTS matematika berkonteks budaya *Nyadran* desa Sonoageng dilakukan oleh dua validator. Berdasarkan hasil analisis data validasi oleh dua validator terhadap 20 butir soal HOTS matematika berkonteks budaya *Nyadran* Desa Sonoageng, dapat disimpulkan bahwa secara keseluruhan kualitas butir soal berada pada kategori sangat valid. Hal ini ditunjukkan oleh dominasi persentase pada kategori sangat valid di hampir seluruh aspek, baik aspek materi, konstruksi, maupun bahasa, dengan persentase berkisar antara 70% hingga 100%. Tidak terdapat butir soal yang masuk dalam kategori tidak valid pada semua indikator yang dinilai. Meskipun masih terdapat beberapa butir soal yang berada pada kategori valid, jumlahnya relatif kecil dan tidak mempengaruhi kelayakan instrumen secara keseluruhan. Dengan demikian, seluruh butir soal yang dikembangkan dinyatakan layak untuk digunakan pada tahap uji coba, karena telah memenuhi kriteria validitas yang baik dari segi kesesuaian materi, ketepatan konstruksi, serta penggunaan bahasa yang sesuai dan komunikatif.

Meskipun secara kuantitatif telah dinyatakan layak, hal ini menunjukkan bahwa proses validasi tidak hanya berhenti pada penilaian angka, tetapi juga mencakup perbaikan substansi soal agar lebih optimal berdasarkan komentar dan saran dari validator. Dengan demikian, revisi yang dilakukan bersifat penyempurnaan untuk meningkatkan kualitas soal, sehingga instrumen yang dihasilkan tidak hanya valid secara statistik, tetapi juga kuat secara konseptual dan kontekstual.

3. Reliabilitas Butir Soal

Reliabilitas butir soal merupakan suatu konsep yang berkaitan dengan tingkat konsistensi dan kestabilan hasil pengukuran yang dihasilkan oleh instrumen tes (Sumintono & Widhiarso, 2015). Dalam konteks ini, reliabilitas menunjukkan sejauh mana butir soal mampu memberikan hasil yang konsisten ketika digunakan untuk mengukur kemampuan yang sama pada kelompok responden.

a. Analisis Reliabilitas Soal dengan *Rasch Model*

Berdasarkan analisis menggunakan *Model Rasch* dengan bantuan software *Ministeps*, reliabilitas instrumen tidak hanya dilihat dari satu indikator, tetapi melalui beberapa aspek, yaitu nilai *Alpha Cronbach*, *Person Reliability*, dan *Item Reliability*. Hasil analisis menunjukkan bahwa nilai *Alpha Cronbach* sebesar 0,52 berada pada kategori rendah, yang mengindikasikan bahwa konsistensi internal antara responden dan butir soal masih belum optimal. Nilai *Person Reliability* sebesar 0,54 yang juga tergolong lemah, sehingga menunjukkan bahwa konsistensi jawaban peserta

didik masih kurang stabil. Kondisi ini dapat disebabkan oleh beberapa faktor, seperti variasi kemampuan peserta didik yang tidak merata, kurangnya pemahaman terhadap konteks soal, serta karakteristik soal HOTS yang menuntut kemampuan berpikir tingkat tinggi sehingga memengaruhi kestabilan respons (Sudijono, 2015; Arikunto, 2018). Di sisi lain, nilai *Item Reliability* sebesar 0,81 berada pada kategori baik, yang menunjukkan bahwa butir soal telah memiliki konsistensi yang cukup tinggi dalam mengukur kemampuan yang dituju. Hal ini berarti bahwa secara konstruksi, kualitas butir soal sudah cukup baik dan mampu membedakan tingkat kemampuan peserta didik.

Hasil dari aspek tingkat kesesuaian butir soal (*item fit*), sebagian besar butir soal telah memenuhi kriteria kesesuaian dengan model Rasch. Hal ini ditunjukkan terutama dari indikator *Outfit Z-Standard (ZSTD)* yang seluruhnya berada dalam rentang yang dapat diterima. Meskipun demikian, terdapat beberapa butir soal yang tidak memenuhi kriteria pada indikator tertentu, seperti *MNSQ* dan *Point Measure Correlation (Pt Mean Corr)*. Butir soal yang secara konsisten tidak memenuhi lebih dari satu indikator, yaitu butir nomor 3, 13, dan 18. Secara umum, instrumen dapat dikatakan cukup baik, namun tetap memerlukan penyempurnaan pada beberapa butir soal. Hasil analisis juga menunjukkan bahwa terdapat 7 butir soal yang tidak masuk dalam kategori fit yaitu butir soal

nomor 9, 10, 12, 14, 15, 19, dan 20 berdasarkan nilai pada item measure, khususnya ditandai dengan munculnya nilai *minimum measure*.

Hasil dari aspek tingkat kesukaran butir soal (*item measure*), diperoleh variasi tingkat kesukaran yang mencakup kategori mudah, sedang, sulit, hingga sangat sulit. Distribusi menunjukkan bahwa proporsi terbesar berada pada kategori sedang (40%), diikuti kategori mudah (35%), sulit (15%), dan sangat sulit (10%). Kondisi ini menunjukkan bahwa sebagian besar butir soal berada pada tingkat kesukaran yang ideal, karena soal kategori sedang mampu memberikan informasi yang paling optimal dalam membedakan kemampuan peserta didik.

Hasil penelitian di atas sejalan dengan temuan bahwa distribusi tingkat kesukaran yang didominasi kategori sedang mampu memberikan pengukuran yang lebih optimal terhadap kemampuan peserta didik (Harbit et al., 2024). Sebaliknya, penelitian lain menunjukkan bahwa distribusi soal yang tidak merata dan didominasi oleh kategori mudah menyebabkan instrumen kurang mampu membedakan kemampuan peserta didik secara akurat (Sukmawati & Kamarudin, 2024). Selain itu, dominasi soal mudah juga mengakibatkan rendahnya daya ukur instrumen karena sebagian besar peserta didik dapat menjawab dengan benar (Aini et al., 2025). Perbedaan penelitian di atas dapat disebabkan oleh

karakteristik peserta didik, konteks soal yang digunakan, serta tujuan pengembangan instrumen yang berbeda.

Hasil dari aspek tingkat abilitas individu (*person measure*), hasil analisis menunjukkan bahwa sebagian besar peserta didik berada pada kategori kemampuan tinggi dan sedang, masing-masing sebesar 48%, sedangkan kategori rendah hanya 4%. Hal ini menunjukkan bahwa secara umum peserta didik telah memiliki kemampuan yang cukup baik dalam menyelesaikan soal HOTS. Temuan ini sejalan dengan penelitian yang menunjukkan bahwa mayoritas siswa berada pada kategori sedang hingga tinggi dalam menyelesaikan soal HOTS (Putri et al., 2020). Namun demikian, terdapat pula penelitian yang menunjukkan bahwa sebagian besar peserta didik masih berada pada kategori rendah dalam menyelesaikan soal HOTS (Wulandari, 2020). Perbedaan tersebut dapat disebabkan oleh beberapa faktor, seperti perbedaan karakteristik peserta didik, kualitas pembelajaran, serta tingkat keterbiasaan peserta didik dalam mengerjakan soal HOTS.

b. Analisis Reliabilitas Soal dengan Tes Teori Klasik

Berdasarkan hasil analisis menggunakan Teori Tes Klasik, dapat disimpulkan bahwa tingkat reliabilitas instrumen yang dikembangkan masih tergolong rendah, dengan nilai *Alpha Cronbach* sebesar 0,54. Hal ini menunjukkan bahwa konsistensi antarbutir soal dalam mengukur kemampuan peserta didik

belum optimal, sehingga masih terdapat variasi respons yang cukup tinggi. Hasil penelitian di atas sejalan dengan temuan bahwa nilai *Alpha Cronbach* berada pada kategori rendah, yang mengindikasikan bahwa konsistensi internal antara responden dan butir soal masih belum optimal (Erfan, 2020). Rendahnya nilai reliabilitas ini dapat dipengaruhi oleh karakteristik butir soal yang berdasarkan analisis Teori Tes Klasik cenderung berada pada kategori mudah. Soal yang terlalu mudah umumnya kurang mampu membedakan kemampuan peserta didik secara optimal, karena sebagian besar siswa dapat menjawab dengan benar (Arifin, 2012). Akibatnya, variasi skor antar peserta didik menjadi kecil, sehingga berdampak pada rendahnya nilai reliabilitas. Kondisi ini mengindikasikan bahwa instrumen masih memerlukan perbaikan, baik dari segi kualitas butir soal maupun kesesuaian dengan konstruk yang diukur, agar mampu menghasilkan pengukuran yang lebih konsisten (Erfan, 2020). Sebaliknya, penelitian lain menunjukkan bahwa nilai *Alpha Cronbach* berada pada kategori sedang yang mengindikasikan bahwa konsistensi antara responden dan butir soal cukup baik (Jumini, 2022).

Selain itu, berdasarkan analisis tingkat kesukaran butir soal, diperoleh bahwa sebagian besar butir soal berada pada kategori mudah, yaitu sebesar 85%, sedangkan kategori sedang sebesar 10%

dan kategori sulit hanya sebesar 5%. Distribusi ini menunjukkan bahwa instrumen telah memiliki variasi tingkat kesukaran, namun belum proporsional karena didominasi oleh soal mudah. Dominasi soal mudah mengindikasikan bahwa sebagian besar peserta didik dapat menjawab soal dengan benar, sehingga instrumen kurang optimal dalam membedakan kemampuan peserta didik secara lebih tajam. Dominasi butir soal pada kategori mudah mengindikasikan bahwa instrumen cenderung kurang mampu membedakan kemampuan peserta didik secara optimal. Soal yang terlalu mudah umumnya dapat dijawab oleh sebagian besar peserta didik, sehingga variasi skor menjadi kecil. Kondisi ini berdampak terhadap nilai reliabilitas instrumen secara keseluruhan. Dengan demikian, dapat disimpulkan bahwa secara umum instrumen yang dikembangkan masih perlu disempurnakan, khususnya dalam meningkatkan reliabilitas dan menyeimbangkan distribusi tingkat kesukaran butir soal. Perbaikan ini penting agar instrumen dapat mengukur kemampuan peserta didik secara lebih akurat dan konsisten.

c. Perbandingan Analisis Reliabilitas Soal dengan *Rasch Model* dan Reliabilitas Soal dengan Tes Teori Klasik

Pada analisis menggunakan *Model Rasch*, reliabilitas ditinjau melalui beberapa indikator, yaitu *Alpha Cronbach*, *Person Reliability*, dan *Item Reliability*. Nilai *Alpha Cronbach* sebesar 0,52 menunjukkan bahwa konsistensi interaksi antara responden dan

butir soal masih tergolong rendah. Nilai *Person Reliability* sebesar 0,54 yang juga berada pada kategori rendah, yang mengindikasikan bahwa konsistensi jawaban peserta didik belum stabil. Namun demikian, nilai *Item Reliability* sebesar 0,81 menunjukkan kategori baik, yang berarti butir soal yang dikembangkan telah memiliki kualitas yang cukup konsisten dalam mengukur kemampuan yang diharapkan.

Jika dibandingkan, kedua pendekatan menunjukkan hasil yang relatif konsisten dalam menilai reliabilitas instrumen, yaitu sama-sama berada pada kategori rendah untuk aspek konsistensi internal. Namun, *Model Rasch* memberikan informasi yang lebih rinci dengan memisahkan antara reliabilitas peserta didik (*Person Reliability*) dan reliabilitas butir soal (*Item Reliability*). Dari hasil tersebut terlihat bahwa kelemahan utama instrumen terletak pada konsistensi respons peserta didik, bukan pada kualitas butir soal, karena *item reliability* sudah tergolong baik.

Hasil analisis tingkat kesukaran butir soal menggunakan *Model Rasch* dan Teori Tes Klasik menunjukkan adanya perbedaan yang cukup signifikan dalam distribusi kategori kesukaran. Pada *Model Rasch*, tingkat kesukaran butir soal diukur menggunakan nilai *logit* yang memperhatikan interaksi antara kemampuan peserta didik dan karakteristik butir soal. Hasil analisis menunjukkan bahwa proporsi terbesar berada pada kategori sedang (40%), diikuti

kategori mudah (35%), sulit (15%), dan sangat sulit (10%). Sebaliknya, berdasarkan analisis menggunakan Teori Tes Klasik yang mengacu pada proporsi jawaban benar, sebagian besar butir soal justru berada pada kategori mudah, yaitu sebesar 85%, sedangkan kategori sedang hanya 10% dan kategori sulit sebesar 5%. Hasil ini menunjukkan bahwa menurut pendekatan Teori Tes Klasik, instrumen cenderung dianggap mudah oleh peserta didik.

Perbedaan hasil ini disebabkan oleh perbedaan dasar pengukuran antara kedua pendekatan tersebut. *Model Rasch* menggunakan pendekatan probabilistik yang mempertimbangkan kemampuan individu dan tingkat kesukaran butir secara simultan, sehingga menghasilkan estimasi yang lebih objektif dan bebas dari ketergantungan pada sampel. Sementara itu, Teori Tes Klasik sangat bergantung pada proporsi jawaban benar peserta didik, sehingga hasilnya dipengaruhi oleh kemampuan kelompok sampel yang digunakan dalam uji coba (Haridanti, 2023).

Akibatnya, suatu butir soal yang tampak mudah dalam Teori Tes Klasik belum tentu benar-benar mudah secara konstruk, karena bisa jadi peserta didik yang menjadi sampel memiliki kemampuan relatif tinggi. Sebaliknya, *Model Rasch* mampu mengidentifikasi bahwa butir tersebut sebenarnya memiliki tingkat kesukaran yang tinggi jika dibandingkan dengan kemampuan individu secara keseluruhan. Dengan demikian, Model Rasch memberikan

informasi yang lebih mendalam dan komprehensif terkait kualitas butir soal, terutama dalam hal distribusi tingkat kesukaran dan kesesuaian dengan kemampuan peserta didik. Namun demikian, hasil dari Teori Tes Klasik tetap memberikan gambaran empiris awal mengenai performa butir soal berdasarkan respons peserta didik.

B. Saran Pemanfaatan, Diseminasi, dan Pengembangan Produk Lebih Lanjut

1. Saran Pemanfaatan Produk

Berdasarkan hasil penelitian dan pengembangan soal HOTS matematika berkonteks budaya Nyadran Desa Sonoageng, peneliti memberikan beberapa saran terkait pemanfaatan instrumen yang telah dikembangkan.

a. Bagi Sekolah

Dengan adanya soal HOTS matematika berkonteks budaya *Nyadran* dapat digunakan sebagai latihan persiapan ujian akhir semester dan tes kemampuan akademik.

b. Bagi Guru

Soal HOTS matematika berkonteks budaya *Nyadran* desa Sonoageng dapat dimanfaatkan sebagai inovasi baru dalam membuat soal evaluasi HOTS berkonteks budaya lokal.

c. Bagi Peserta Didik

Soal HOTS matematika berkonteks budaya *Nyadran* desa Sonoageng dapat dimanfaatkan untuk melatih kemampuan berpikir

tingkat tinggi, memudahkan peserta didik dalam mengerjakan soal-soal kemampuan berpikir tingkat tinggi.

2. Saran Diseminasi Produk

Produk pengembangan soal HOTS matematika berkonteks budaya Nyadran desa Sonoageng dapat digunakan untuk semua sekolah SMA/MA kelas XI di kabupaten Nganjuk, akan tetapi untuk penyebaran produk pengembangan tetap harus memperhatikan dan mempertimbangkan peraturan sekolah, karena terdapat beberapa sekolah tidak memperbolehkan siswa membawa alat elektronik. Selain itu, diseminasi produk ini juga disarankan dilakukan melalui forum Musyawarah Guru Mata Pelajaran (MGMP) Matematika. Melalui MGMP, produk yang dikembangkan dapat diperkenalkan, didiskusikan, serta diimplementasikan secara lebih luas oleh para guru, sehingga kebermanfaatannya dapat dirasakan secara optimal dalam meningkatkan kemampuan berpikir tingkat tinggi (HOTS) peserta didik.

3. Saran Pengembangan Produk Lebih Lanjut

Adapun saran pengembangan produk lebih lanjut adalah sebagai berikut:

- a. Bagi peneliti selanjutnya yang ingin mengembangkan produk ini, disarankan untuk menggunakan bentuk soal yang lebih variatif, seperti pilihan ganda kompleks, menjodohkan, isian singkat, serta soal uraian. Variasi bentuk soal ini bertujuan untuk meminimalisasi kemungkinan peserta didik dalam menebak jawaban, sehingga kemampuan berpikir tingkat tinggi dapat diukur secara lebih akurat.

- b. Peneliti selanjutnya juga disarankan untuk melakukan uji coba produk dengan jumlah sampel yang lebih besar dan beragam, misalnya dengan melibatkan seluruh peserta didik kelas XI pada sekolah yang bersangkutan atau bahkan pada beberapa sekolah yang berbeda. Hal ini bertujuan untuk meningkatkan keakuratan hasil analisis serta memperkuat generalisasi kualitas instrumen yang dikembangkan.
- c. Peneliti selanjutnya yang ingin mengembangkan produk ini disarankan untuk menggunakan seluruh elemen serta materi yang terdapat pada fase F dan F+. Hal ini penting agar cakupan materi yang dikembangkan menjadi lebih komprehensif dan mampu merepresentasikan kemampuan peserta didik secara lebih menyeluruh, khususnya dalam mengukur keterampilan berpikir tingkat tinggi. Selain itu, dengan mengintegrasikan seluruh elemen pada kedua fase tersebut, produk yang dihasilkan diharapkan memiliki variasi konteks dan tingkat kesulitan yang lebih beragam, sehingga dapat meningkatkan kualitas instrumen serta relevansinya terhadap kurikulum yang berlaku.
- d. Pengembangan produk selanjutnya dapat difokuskan pada perbaikan butir soal yang belum memenuhi kriteria, terutama dari segi reliabilitas dan tingkat kesukaran. Peneliti juga dapat menyeimbangkan distribusi tingkat kesukaran agar instrumen yang

dihasilkan lebih proporsional dan mampu mengukur kemampuan peserta didik secara menyeluruh.

- e. Pengembangan lebih lanjut juga dapat dilakukan dengan memperluas konteks budaya yang digunakan, tidak hanya terbatas pada budaya *Nyadran* Desa Sonoageng, tetapi juga budaya lokal lainnya. Hal ini bertujuan agar produk yang dikembangkan menjadi lebih variatif, kontekstual, serta memiliki daya jangkau dan kebermanfaatan yang lebih luas dalam dunia pendidikan.