



Proses Respon sebagai Bukti Validitas dalam Pengembangan Tes di Bidang Pendidikan

Firmanto Adi Nurcahyo

Fakultas Psikologi Universitas Pelita Harapan Surabaya

Abstrak

Tes merupakan suatu hal yang penting dalam bidang pendidikan. Sebuah tes yang baik harus dapat menghasilkan skor yang akurat dan merefleksikan individu yang sebenarnya. Untuk itu diperlukan bukti-bukti yang mendukung keakuratan interpretasi serta penggunaan suatu tes. Validitas bergantung pada sejauhmana bukti dan teori mendukung interpretasi dari skor yang diperoleh dari suatu tes. Tulisan ini bertujuan untuk memberikan gambaran mengenai konsep validitas dalam pengembangan tes di bidang pendidikan. Secara lebih khusus, pemaparan berdasarkan studi literatur mengenai proses respon sebagai bukti validitas menjadi fokus dari tulisan ini.

Kata kunci: pengembangan tes; validitas; proses respon

A. PENDAHULUAN

Pengukuran dalam pendidikan merupakan hal yang penting. Hal ini dikarenakan hasil dari pengukuran dalam bidang pendidikan seringkali menjadi dasar dalam pengambilan suatu keputusan. Keputusan yang diambil tersebut dapat menyebabkan individu untuk naik kelas, lulus sekolah, diterima pada suatu pendidikan lanjutan, diterima dalam suatu pekerjaan, ataupun promosi jabatan. Secara umum hasil dari suatu pengukuran dalam pendidikan bisa mengubah kehidupan seseorang (McCoubrie, 2004).

Pengukuran dalam bidang pendidikan pada umumnya dilakukan dengan tes. Berdasarkan definisi yang diajukan para ahli, Azwar (2005) menyimpulkan bahwa tes adalah prosedur yang sistematis, berisi sampel perilaku, dan mengukur perilaku. Ini berarti terdapat prosedur tertentu dalam penyajian dan pemberian nilai dalam tes. Selain itu, pertanyaan-pertanyaan yang diajukan dalam tes harus merupakan representasi dari domain perilaku yang diukur.

Tes dapat dibedakan menjadi tes yang mengungkap atribut kognitif dan non kognitif (Azwar, 2005). Tes kognitif mengungkap apa yang mampu dilakukan



individu dan seberapa baik individu melakukannya. Tes intelegensi dan tes merupakan contoh tes kognitif. Tes non kognitif mengungkap kecenderungan perilaku individu dalam situasi-situasi tertentu. Tes minat serta kepribadian merupakan contoh dari tes non kognitif.

Berbagai tes telah banyak dikembangkan dalam bidang pendidikan. Tes evaluasi belajar, ujian nasional, Tes Potensi Akademik (TPA), serta tes intelegensi merupakan contoh tes kognitif yang banyak dipakai saat ini. Berbagai tes tersebut bertujuan untuk menguji kemampuan berpikir individu. Di sisi lain, terdapat juga berbagai tes untuk mengukur kemampuan non-kognitif dalam bidang pendidikan seperti misalnya tes minat vokasional, gaya belajar, serta sikap terhadap suatu matapelajaran.

Pengembangan tes dalam bidang pendidikan perlu terus dilakukan. Pengembangan kurikulum menuntut guru untuk mengembangkan tes evaluasi belajar sesuai dengan kurikulum yang baru. Tes Potensi Akademik selalu membutuhkan pembaharuan soal. Perkembangan dalam teori intelegensi akan diikuti dengan usaha pengembangan tes untuk mengukurnya.

Sebuah tes yang baik harus dapat menghasilkan skor yang akurat dan merefleksikan individu yang sebenarnya. Untuk bisa menghasilkan skor yang akurat, sebuah tes harus memiliki properti psikometri yang baik. Properti psikometri yang penting dalam pengembangan tes adalah validitas dan reliabilitas. Kedua hal tersebut merupakan elemen fundamental dalam evaluasi suatu tes (Tavakol & Dennick, 2011).

Tulisan ini bertujuan untuk memberikan gambaran mengenai konsep validitas dalam pengembangan tes. Konsep validitas didasarkan pada *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). Secara lebih khusus pemaparan mengenai proses respon sebagai bukti validitas akan menjadi fokus tulisan ini.

B. TINJAUAN TEORETIS

B.1 Pengembangan Tes

Pengembangan tes dapat dilakukan melalui adaptasi. Pada proses adaptasi, suatu



tes yang dibangun dari suatu bahasa dan budaya tertentu, diadaptasikan pada budaya lain. Hal ini bertujuan agar tes tersebut sesuai untuk dipakai pada kondisi budaya yang dituju.

Dalam melakukan adaptasi, adanya informasi mengenai kualitas psikometris dari tes yang hendak diadaptasikan merupakan hal yang penting. Kualitas psikometris dari tes yang hendak diadaptasikan harus terbukti bagus. Kualitas psikometris dari tes pun nantinya akan diuji sebagai bagian dari proses adaptasi.

Pengembangan tes juga bisa dilakukan dengan menyusun instrumen yang baru. Penyusunan tes baru dapat dilakukan karena belum adanya tes yang bisa mengukur konstruk yang diinginkan. Selain itu, penyusunan tes yang baru juga umum dilakukan sebagai usaha untuk mengatasi permasalahan yang ada pada tes sebelumnya.

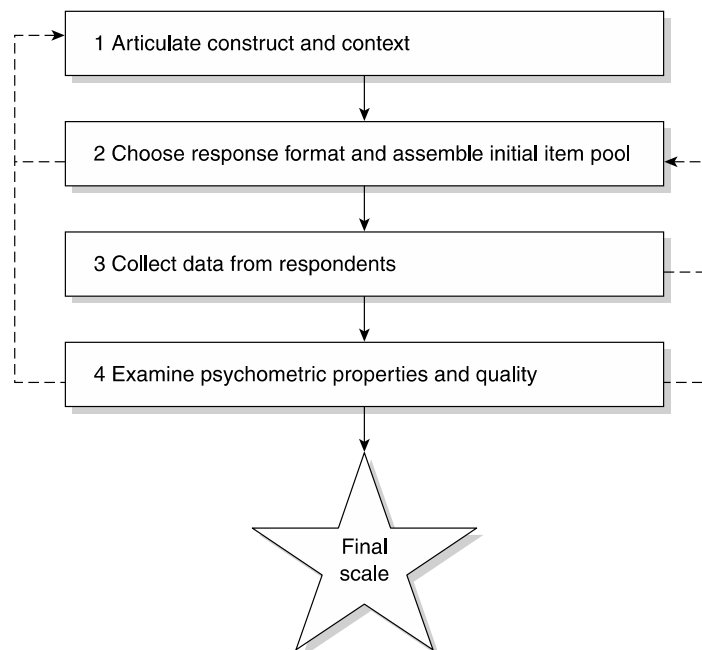
Furr (2011) mengungkapkan bahwa penyusunan tes melalui empat tahap (Gambar 1). Keempat tahapan tersebut adalah mengartikulasi konstruk dan konteks, memilih format respon dan membangun aitem awal, mengumpulkan data dari partisipan, dan menguji properti psikometri. Masing-masing tahapan dapat dijelaskan sebagai berikut.

- a. Tahap pertama dalam pengembangan tes adalah mengartikan variabel yang hendak diukur serta mendiferensiasi variabel tersebut dari konstruk lain. Pada tahap ini penyusun tes perlu memahami konteks tes akan digunakan. Konteks tersebut meliputi target populasi dan konteks administrasi. Target populasi akan memberi arah dalam penulisan aitem. Format respon, aitem, dan instruksi untuk orang dewasa tentunya berbeda dengan anak-anak.
- b. Tahap kedua merupakan usaha untuk menulis aitem yang relevan dengan konstruk yang diukur. Hal ini sangat terkait dengan jumlah konstruk yang diukur, panjang pendeknya tes, serta kejelasan definisi dari konstruk yang diukur. Pada tahap ini, aitem dipertimbangkan dalam hal kejelasan makna serta relevansinya secara konseptual. Aitem yang dianggap kurang memenuhi syarat dimungkinkan untuk dibuang atau direvisi.
- c. Tahap ketiga adalah tahap aitem diberikan pada partisipan yang mewakili populasi sasaran. Tahap ini memiliki tujuan mengungkap permasalahan yang



ada berdasarkan umpan balik dari partisipan. Selain itu, tahap ini juga menghasilkan data yang dipergunakan untuk evaluasi properti dan kualitas psikometri dari aitem.

- d. Tahap keempat merupakan tahapan pemerolehan reliabilitas dan validitas. Jika hasil analisis mengungkapkan kualitas psikometri yang kurang baik, penyusun tes dapat kembali menulis aitem. Proses ini akan menghasilkan bentuk tes akhir dengan kualitas psikometri yang baik.



Gambar 1. Tahapan penyusunan tes (Furr, 2011)

Pengembangan tes baik berupa adaptasi maupun penyusunan tes yang baru pasti melalui proses pemerolehan bukti psikometri. Validitas merupakan salah satu properti psikometri yang perlu untuk dikaji dalam pengembangan suatu tes. Oleh karena itu, pembuktian validitas merupakan suatu proses yang penting.

B.2 Validitas

Validitas merujuk pada sejauh mana bukti dan teori mendukung interpretasi dari skor tes (AERA et al., 2014). Konsep validitas tersebut menjadi rujukan dari para ahli dalam memberikan definisi terhadap validitas. Hubley & Zumbo (2011)



mengungkapkan bahwa validitas terkait tentang inferensi, interpretasi, aksi, serta keputusan yang dibuat berdasarkan skor tes. Furr (2011) menyatakan bahwa validitas mengarah pada sejauhmana bukti dan teori mendukung interpretasi dari skor yang diperoleh dari tes. Pemahaman ini mengarah pada beberapa implikasi yakni validitas menekankan pada intepretasi skor tes, bukan tes itu sendiri. Melakukan validasi interpretasi skor tes berarti melakukan evaluasi pernyataan-pernyataan yang masuk akal berdasar skor (Kane, 2013).

Konsep validitas menyatakan bahwa sebuah atribut yang dibangun secara teoretis itu ada, dan pengukuran dari atribut tersebut dapat dilakukan dengan sebuah tes karena skor tes disebabkan oleh variasi dalam atribut (Borsboom, Mellenbergh, & van Heerden, 2004). Sebuah tes dikatakan valid untuk mengukur sebuah atribut jika variasi dalam atribut menyebabkan variasi pada skor tes.

Validitas adalah masalah tingkatan sehingga validitas harus ditimbang berdasarkan segi kuat lemahnya bukti yang ada (Cook & Beckman, 2006; Furr 2011; Kane, 2013). Tingkatan tersebut bergantung pada sejauhmana bukti dan teori mendukung interpretasi dari skor yang diperoleh dari tes (AERA et al., 2014). Validitas yang tinggi ditunjukkan dengan interpretasi yang didukung oleh bukti-bukti yang tepat. Sebaliknya interpretasi yang tidak didukung atau bertentangan dengan bukti-bukti yang ada menunjukkan validitas yang rendah.

Validitas dapat berubah dari waktu ke waktu selama adanya perkembangan interpretasi dan terkumpulnya bukti-bukti baru (Kane, 2013). Ini berarti validitas dapat menurun jika bukti-bukti baru menunjukkan adanya keraguan terhadap interpretasi skor yang telah ada sebelumnya. Ini mengindikasikan bahwa proses pemerolehan bukti validitas tidak selesai dalam suatu waktu saja.

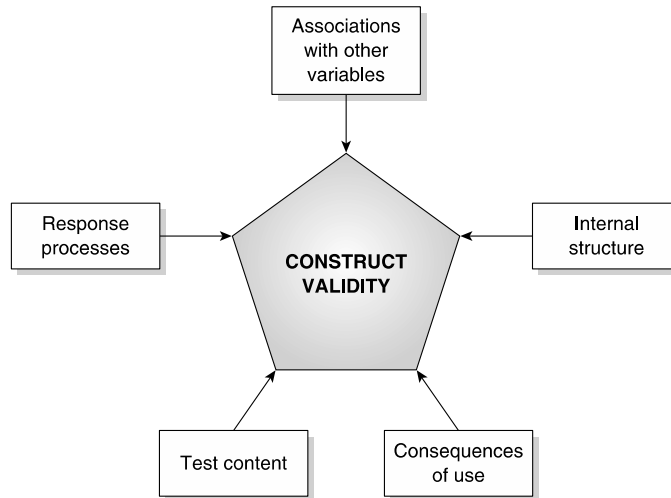
B.3 Bukti Validitas

American Education Research Association, American Psychological Association, dan National Council on Measurement in Education (2014) mengungkapkan lima jenis bukti yang relevan dengan validitas yakni bukti isi tes, struktur internal, proses psikologis dalam merespon tes, konsekuensi penggunaan tes, serta keterkaitan tes





dengan variabel lain. Furr (2011) mengilustrasikan kelima bukti validitas tersebut terpusat pada validitas konstruk (Gambar 2).



Gambar 2. Sumber bukti validitas (Furr, 2011)

Validitas isi menekankan bahwa isi tes harus menggambarkan aspek-aspek penting dari konstruk psikologis yang hendak diukur (Furr, 2011). Ketidakvalidan terjadi ketika tes berisi konstruk yang tidak relevan (*construct irrelevant content*). Bukti yang kuat dari validitas isi diperoleh melalui pemeriksaan serta persetujuan dari para ahli. Kriteria ahli meliputi orang yang berpengalaman serta memiliki kompetensi dalam konstruk/variabel yang ingin dibuat sebagai tes (Rico et al., 2012).

Bukti validitas terkait dengan struktur internal atau dimensionalitas dari tes. Jika suatu tes diinterpretasikan sebagai pengukur konstruk tertentu, maka struktur aktual yang ada harus sesuai dengan struktur berdasarkan konstruk teoritis (Furr, 2011). Jika suatu tes dimaksudkan untuk menggambarkan konstruk tunggal, maka struktur internal yang diharapkan bersifat unidimensi.

Bukti konsekuensi merujuk pada dampak, manfaat, serta apa yang diharapkan dari sebuah pengukuran (Cook & Lineberry, 2016). Administrasi pelaksanaan tes, analisis dan interpretasi skor, serta pengambilan keputusan memiliki dampak langsung terhadap partisipan yang dikenai tes serta orang lain. Dampak inilah yang harus





dievaluasi untuk memastikan bahwa manfaat yang diperoleh dapat mengatasi kerugian yang mungkin ditimbulkan.

Bukti validitas juga ditunjukkan melalui asosiasi antara skor tes dan hasil pengukuran dari variabel psikologis lain (Furr, 2011). Dalam hal ini, validitas didasarkan pada kesesuaian antara sebuah tes dengan tes lain secara teoretis dan aktual. Semakin sesuai hasil yang diperoleh, semakin tinggi keyakinan bahwa skor tes merepresentasikan suatu konstruk yang diukur.

B.4 Proses Respon sebagai Bukti Validitas

Analisis teoritis dan empiris dari proses respon partisipan dapat memberikan bukti mengenai kesesuaian antara konstruk dengan detail kinerja atau respon yang diberikan oleh partisipan (AERA et al., 2014). Dengan kata lain bukti validitas dapat ditunjukkan dari kesesuaian antara proses psikologis yang secara aktual mempengaruhi respon dan proses yang seharusnya berpengaruh terhadap respon (Furr, 2011). Suatu tes merupakan pengukur yang valid dari suatu konstruk jika konstruk yang dimaksud benar-benar mempengaruhi respon partisipan terhadap tes.

Terdapat suatu proses yang melatarbelakangi individu dalam memberikan tanggapan terhadap aitem-aitem yang akhirnya terwujud dalam bentuk skor tes. Proses yang dimaksud adalah mekanisme teoritis yang mendasari respons aitem, termasuk di dalamnya adalah proses, strategi, dan pengetahuan yang diterapkan partisipan dalam memberikan respon terhadap aitem (Embretson, 2016). Borsboom, Mellenbergh, & van Heerden (2004) mengungkapkan bahwa penyelidikan terhadap validasi tes bertujuan untuk menawarkan penjelasan teoritis mengenai proses yang mengarah pada hasil pengukuran. Dengan demikian perhatian terhadap validitas harus dimulai dari teori yang jelas mengenai proses psikologis yang mendasari respon dari partisipan.

Bukti validitas proses respon perlu diwujudkan secara nyata. Beberapa indikasi untuk memperoleh bukti proses respon misalnya mempertanyakan pada partisipan mengenai strategi dalam merespon aitem, membuat catatan untuk memonitor perkembangan respon, mendokumentasikan aspek-aspek kinerja seperti pergerakan mata serta waktu reaksi (Padilla & Benítez, 2014). Inferensi mengenai proses yang



terlibat pada respon dapat juga dikembangkan dengan melakukan analisa keterkaitan antar bagian-bagian tes serta antara tes dan variabel lain (AERA et al., 2014).

B.5 Bukti Proses Respon pada Aitem Analogi Verbal

Analogi Verbal merupakan tipe aitem untuk mengukur kecerdasan, dan merupakan salah satu tipe aitem pada Tes Potensi Akademik (Embretson, 2016). Pada model tes tersebut, partisipan diminta untuk memahami hubungan antara dua kata. Keterhubungan pada kedua kata tersebut lalu dianalogikan pada kedua kata yang lain.

Whitely (1977) melakukan studi untuk mengetahui bagaimana proses respon partisipan terhadap aitem analogi. Partisipan diminta untuk mengerjakan 60 soal analogi verbal. Setelah itu partisipan diminta untuk mengelompokkan soal-soal tersebut berdasarkan tipe keterhubungannya.

Hasil penelitian Whitely (1977) menunjukkan adanya 8 pengelompokan tipe relasi kata (Tabel 1). Empat tipe relasi (*Opposites*, *Similarities*, *Class Membership*, dan *Class Naming*) berisikan aitem-aitem yang terkait dengan arti dari kata. Tiga tipe relasi (*Functional*, *Conversion*, dan *Quantitative*) berhubungan dengan properti dari objek. Tipe relasi lain yakni *Word Pattern* meliputi relasi nonsemantik berdasarkan ciri-ciri tertentu misalnya rima pada kata.

Hasil penelitian Whitely (1977) menunjukkan bahwa pola-pola relasi yang terbentuk sesuai dengan apa yang dimaksudkan dalam penyusunan aitem-aitem analogi verbal. Hal ini menguatkan bahwa proses respon yang diberikan oleh partisipan telah seperti yang diharapkan. Dalam hal ini partisipan mencari pola keterhubungan antar kata. Pola keterhubungan yang telah diperoleh lalu dipergunakan untuk menjawab kata yang dipertanyakan.

Tabel 1. Tipe relasi kata (diadaptasi dari Embretson, 2016)

Relasi	Definisi	Contoh
<i>Opposites</i>	Lawan kata	Konkret : abstrak = gasal : genap
<i>Similarities</i>	Keserupaan kata	Konkaf : cekung = konveks : cembung
<i>Class Membership</i>	Kesamaan kategori	Sawi : bayam = mangga : pepaya
<i>Class Naming</i>	Kelas & anggota	Hewan : gajah = mobil : sedan





<i>Functional</i>	Objek dengan fungsinya	Gigi : kunyah = jari : pegang
<i>Conversion</i>	Objek yang berubah ke bentuk lain	Ulat : kupu-kupu = bayi : remaja
<i>Quantitative</i>	Objek terkait dengan ukuran atau kuantitas	Detik : menit = bulan : tahun
<i>Word Pattern</i>	Kata berima	Pot : got = amis : manis

B.6 Bukti Proses Respon melalui Wawancara Kognitif

Proses respon yang dilakukan individu dalam menjawab suatu tes atau kuisisioner tidak bersifat otomatis. Individu menggunakan informasi-informasi yang tersimpan dan tersedia dalam memori untuk menjawab (Conrad & Blair, 2009). Informasi-informasi itu dapat digali melalui wawancara kognitif.

Beatty and Willis (2007) mendeskripsikan wawancara kognitif sebagai administrasi draft pertanyaan-pertanyaan survei bersamaan dengan mengumpulkan informasi tambahan tentang respon survei, yang dimaksudkan untuk mengevaluasi kualitas respon atau untuk membantu menentukan apakah pertanyaan dapat menghasilkan informasi seperti yang dimaksudkan oleh peneliti. Tujuan dari wawancara kognitif adalah untuk mengakses proses kognitif individu sebagai bukti validitas proses respon dari individu. Wawancara kognitif seringkali dipergunakan dalam riset survei dimana peneliti bermaksud memahami proses kognitif yang terjadi pada saat partisipan menjawab pertanyaan-pertanyaan survei (Castillo, Padilla, Gomez, & Andres, 2010).

Istilah teknis pertanyaan dalam wawancara kognitif adalah “*probes*” atau “*follow-up probes*” (Padilla & Benítez, 2014). Terdapat dua strategi dalam mengembangkan “*probes*” saat melakukan wawancara kognitif. Strategi pertama adalah melalui metode *think a loud*. Inti dari metode ini adalah verbalisasi dari pemikiran-pemikiran dari responden saat mereka merespon terhadap tes. Strategi kedua adalah metode berdasar “*probe*”. Metode ini dilakukan dengan mengembangkan “*probes*” untuk area-area khusus dari setiap aitem. Dalam hal ini, “*probes*” dikembangkan atas dasar ciri-ciri atau elemen-elemen dari aitem yang dirasa peneliti cukup problematik terkait



oleh partisipan atau kemungkinan suatu kelompok partisipan menginterpretasikan aitem secara berbeda.

Wawancara kognitif digunakan untuk mengidentifikasi permasalahan dalam pengembangan tes dengan menanyakan sekelompok kecil partisipan untuk melaporkan secara verbal pemikiran mereka ketika menjawab pertanyaan (Conrad & Blair, 2009). Dalam hal ini partisipan diminta untuk mengerjakan tes sesuai dengan prosedur administrasi, lalu dilanjutkan dengan pelaksanaan wawancara kognitif (Padilla, 2014). Wawancara tersebut dilakukan untuk menggali jawaban-jawaban yang telah diberikan partisipan.

Castillo-Diaz dan Padilla (2013) menunjukkan bagaimana wawancara kognitif dilakukan sebagai bukti proses respon. Dalam penelitian tersebut partisipan diminta untuk memberikan respon terhadap APGAR, tes yang ditujukan untuk mengukur tingkat kepuasan individu terhadap dukungan keluarga. Aitem dalam APGAR misalnya adalah “Apakah Anda merasa puas dengan bantuan yang Anda terima dari keluarga ketika Anda mengalami masalah?” serta “Apakah Anda merasa dicintai oleh keluarga Anda?”. Partisipan diminta untuk merespon salah satu pilihan yakni “selalu, kadang-kadang, tidak pernah”. Setelah itu, wawancara kognitif dilakukan terhadap partisipan. Wawancara kognitif dilakukan untuk mendapatkan argumen dari partisipan mengenai konsep keluarga, anggota keluarga, serta tingkatan respon jawaban. Penggalan data melalui wawancara dilakukan seperti terlihat pada Tabel 2.

Hasil wawancara Castillo-Diaz dan Padilla (2013) menunjukkan adanya tiga informasi. Partisipan menjawab aitem-aitem APGAR berdasarkan pemahaman terhadap dukungan keluarga serta memiliki pengertian konsep keluarga yang konstan saat menjawab aitem-aitem APGAR. Berkaitan dengan anggota keluarga, partisipan tidak hanya memasukkan keluarga dekat tetapi juga kerabat serta teman dalam merespon aitem-aitem APGAR. Hasil wawancara terhadap tingkatan respon menunjukkan bahwa konstak dukungan keluarga dapat direspon secara bertingkat mulai dari tidak pernah hingga selalu. Hasil-hasil tersebut menunjukkan bahwa APGAR dapat digunakan untuk mengukur konstak dukungan keluarga. Namun



demikian, APGAR kurang sesuai digunakan dalam pengukuran dukungan keluarga dengan batasan keluarga dekat saja.

Tabel 2. *Probes* dalam wawancara kognitif (Diadaptasi dari Castillo-Diaz dan Padilla (2013))

Argumen	<i>Probes</i>
Konsep keluarga	Apa yang Anda pikirkan saat menjawab pertanyaan-pertanyaan dalam kuisioner? Apa yang Anda pahami dengan “menerima bantuan dari keluarga ketika Anda mengalami permasalahan?”
Anggota keluarga	Ada berapa anggota keluarga dekat Anda? Siapakah anggota keluarga Anda yang membantu Anda ketika mengalami permasalahan?
Tingkatan respon	Dalam pertanyaan “Apakah Anda merasa dicintai oleh keluarga Anda?” Anda menjawab kadang-kadang. Dalam situasi seperti apa jawaban Anda dapat menjadi “selalu” atau “tidak pernah”?

C. KESIMPULAN

Pengembangan tes dalam bidang pendidikan akan terus dilakukan. Hal ini didorong oleh adanya kebutuhan tes-tes yang sesuai dengan perkembangan jaman. Selain itu, adanya perkembangan teori-teori baru juga pasti diikuti dengan kebutuhan pengembangan tes yang sesuai dengan teori tersebut.

Pengembangan tes tidak bisa lepas dari properti psikometrinya. Salah satu properti psikometri yang penting adalah validitas. Untuk itu, usaha pembuktian validitas perlu dilakukan dalam proses pengembangan tes. Validitas yang baik ditunjukkan dengan banyaknya bukti yang mendukung interpretasi dari skor suatu tes.

Bukti validitas proses respon merupakan bukti yang cukup penting dalam pengembangan tes. Ini berkaitan dengan proses bagaimana partisipan memberikan respon terhadap aitem. Bukti validitas proses respon ditunjukkan dari kesesuaian antara proses psikologis yang diharapkan dan yang secara aktual terjadi, yang berpengaruh pada respon terhadap aitem-aitem tes.

Wawancara kognitif merupakan salah satu strategi yang bisa digunakan untuk memperoleh bukti validitas proses respon. Melalui wawancara kognitif penyusun tes



dapat memperoleh informasi bagaimana respon partisipan dalam menjawab aitem-aitem tes. Jika proses respon yang diberikan partisipan tidak sesuai dengan teori atau apa yang diharapkan, penyusun tes dapat mengkaji ulang aitem-aitem tes yang telah dibuat. Hal ini dilakukan untuk menguatkan bahwa aitem-aitem yang disusun dapat direspon sesuai dengan maksud dari aitem tersebut dibuat.

D. DAFTAR PUSTAKA

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Azwar, S. (2005). *Tes Prestasi*. Yogyakarta: Pustaka Pelajar.
- Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, *71*(2), 287–311. <https://doi.org/10.1093/poq/nfm006>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Castillo, M., Padilla, J., Gomez, J., & Andres, A. (2010). A productivity map of cognitive pretest methods for improving survey questions. *Psicothema*, *(22)*, 475–481.
- Castillo-Díaz, M., & Padilla, J.-L. (2013). How Cognitive Interviewing can Provide Validity Evidence of the Response Processes to Scale Items. *Social Indicators Research*, *114*(3), 963–975. <https://doi.org/10.1007/s11205-012-0184-8>
- Conrad, F. G., & Blair, J. (2009). Sources of Error in Cognitive Interviews. *Public Opinion Quarterly*, *73*(1), 32–55. <https://doi.org/10.1093/poq/nfp013>
- Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine*, *119*(2), 166.e7-166.e16. <https://doi.org/10.1016/j.amjmed.2005.10.036>



- Cook, D. A., & Lineberry, M. (2016). Consequences Validity Evidence: Evaluating the Impact of Educational Assessments. *Academic Medicine*, 91(6), 785–795. <https://doi.org/10.1097/ACM.0000000000001114>
- Embretson, S. E. (2016). Understanding Examinees’ Responses to Items: Implications for Measurement. *Educational Measurement: Issues and Practice*, 35(3), 6–22.
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Los Angeles ; London: SAGE.
- Hubley, A. M., & Zumbo, B. D. (2011). Validity and the Consequences of Test Interpretation and Use. *Social Indicators Research*, 103(2), 219–230. <https://doi.org/10.1007/s11205-011-9843-4>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. *Medical Teacher*, 26(8), 709–712.
- Padilla, J.-L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136–144.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach’s alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Whitely, S. E. (1977). Relationships in analogy items: A semantic component of a psychometric task. *Educational and Psychological Measurement*, 37(3), 725–739.

