

TERJEMAHAN

RELIABILITY

*Diajukan untuk memenuhi salah satu tugas mata kuliah "Evaluasi Pembelajaran
B. Arab"*

Dosen :

Drs. Wagino Hamid Hamdani

Dra. Nunung Nursyamsiah, M.Pd



Disusun Oleh:

Dian Rachdiana 0700248

Hani Pursita 0700387

Kiki Amalia 0706047

JURUSAN PENDIDIKAN BAHASA ARAB

FAKULTAS PENDIDIKAN BAHASA DAN SENI

UNIVERSITAS PENDIDIKAN INDONESIA

2009

BAGIAN 7

KEANDALAN

Jika tes tidak memberi arti di luar diri mereka sendiri, mereka tidak akan menjadi lebih penting atau lebih perhatian. Sangat sedikit perilaku di mana pendidik "tertarik, termasuk tes pertunjukan, hanya terjadi dalam satu kali atau sangat terbatas keadaan. Ketika kita menguji, kita tertarik untuk menyamaratakan apa yang kita lihat hari ini di bawah satu set kondisi atau kesempatan lain. (Sebagai contoh jika kita tidak dapat menyamaratakan kemampuan membaca Billy yang diamati selama pengujian ke dalam situasi kelas maka data uji itu kecil atau bernilai). Untuk lebih luas, kita dapat menyamaratakan dari seperangkat fakta-fakta dari pengamatan (sebuah tes, sebagai contoh), pengamatan itu dinamakan keandalan.

Reliabilitas adalah pertimbangan utama dalam evaluasi karakter psikometrik dari skala tes. Sebagai contoh, ketika kita memberi tes pengelolaan kepada seseorang secara individual, kita dapat menyamaratakan hasil dengan tiga cara yang berbeda. Kita dapat mengambil angka ujian dari penguji lain, dan hasilnya akan sama; kita tidak dapat menggunakan perilaku yang menarik perhatian hanya dengan berupa catatan para penguji. Kita juga dapat mengambil dari perilaku kita lihat pada hari ini yang terlihat pada keesokan hari (minggu depan), jika kita melakukan tes lagi; perilaku yang stabil biasanya menarik dalam perbaikan pendidikan. Juga, kita ingin sedikit berasumsi bahwa tes pertanyaan yang asedikit berbeda akan memberikan kita hasil yang mirip; kita akan dapat menyamaratakan item tes lain yang mirip. Dengan demikian, ada tiga jenis reliabilitas. Reliabilitas untuk menyamakan perbedaan angka, yang disebut reabilitas tingkat luar atau angka luar. Reabiliti untuk menyamaratakan waktu yang berbeda disebut reabiliti stabil tes atau mengulang tes. Reabiliti untuk menyamaratakan item tes lain yang disebut reabiliti rumus alternatif atau reabiliti kemantapan internal. Jika skor tes dapat diandalkan dan diuji sifat atau perilaku yang stabil, maka kita dapat menggeneralisasikan skor. Untuk lebih luas, nilai seseorang tidak dapat digeneralisir untuk situasi yang lain, maka angka itu tidak dapat diandalkan. Dalam pendidikan dan psikologi, kita menginginkan tes yang dapat diandalkan.

Reliabilitas dapat digunakan untuk menggambarkan bagaimana seseorang berkeinginan untuk menggeneralisasi pengamatan dan tes. Anggaplah bahwa Nyonya Amig tertarik dalam pengujian siswa TK di atas dan huruf dari abjad Inggris. Ia bisa menilai dumaian-semua lima puluh dua huruf besar dan huruf kecil-atau dia bisa sampel dari domain. Sebagai contoh, ia bisa meminta masing-masing

dari setiap murid untuk ikut menamai surat-surat berikut: A, h, j, L, q, r, R, u, V, w. Dia ingin mengasumsikan bahwa sampel surat lain (misalnya, lg, E, k, m, s, T, U, v, w, z) akan dapat menunjukkan angka yang samadari siswanya. Khususnya setiap akjan mengambil skor yang sama pada setiap sampel dari domain atau pada semua item dalam domain. Jadi, dia ingin menyamaratakan dari sampel ke item yang lainnya dalam domain di mana sampel tersebut drai gambaran sampel. Bagaimana pun, Nyonya Amig ingin lebih dari itu. Anggaplah bahwa ia menetes murid-muridnya pada Senin pagi pukul 9.30. Ia ingin berasumsi bahwa siswa akan memperoleh skor yang sama jika mereka diuji Seni pukul 1:45. Ada domain yang sama dengan item domai. Setiap satu kesempatan ini merupakan contoh dari waktu domain-semua waktu. Nyonya Amig ingin menggeneralisasi hasil hasil pada satu sampel dari waktu ke domain. Pengetahuan murid-muridnyatentang alfabet tidak akan begitu berguna, jika mereka tahu hanya surat-surat pada hari Senin di pukul 9:30.

Cara mudah untuk memikirkan keandalan adalah untuk memikirkan setiap pengukuran (semua diperoleh skor) sebagai terdiri dari dua bagian: nilai yang benar dan nilai yang salah. Kesalahan adalah tidaksesuaiannya dengan skor sejati dan pada dasarnya adalah acak. Kesalahan adalah gagasan terbaik sebagai kurangnya generalisasi bahwa hasil dari untuk mendapatkan perwakilan sampel dari domain. Sebagai contoh, sebuah sampel yang terdiri dari huruf alfabet yang terdiri dari A, B, C, D dan E mungkin akan jauh lebih mudah daripada yang lain dari contoh surat. Sebuah sampel sistematis akan mudah memompa angka Nyonya Amigs diterima oleh siswa. Demikian pula, sebuah sampel dibuat atas tiang surat sulit mungkin akan menurunkan nilai yang diterima oleh murid-muridnya. Jadi, kesalahan-kegagalan untuk memilih sampel yang cocok dengan item-dapat meningkatkan atau menurunkan skor. Rata-rata dari kesalahan dalam jangka panjang adalah sama dengan panjang nol. Dalam jangka panjang, contoh keseimbangan keluar. Skor yang benar adalah skor yang diperoleh ketika seluruh domain dinilai. Mereka juga merupakan rata-rata angka pada semua sampel yang mungkin dari domain. Dalam prestasi murid itu kadang-kadang mungkin untuk menilai seluruh domain. Biasanya, bagaimanapun, penilaian semacam itu tidak mungkin. Seperti yang Anda ingat dari pembahasan asumsi yang mendasari psychological penilaian, pengukuran error selalu hadir. Pertanyaan penting adalah, Berapa banyak kesalahan yang melekat pada nilai tertentu? Sayangnya, jawaban langsung untuk pertanyaan ini tidak tersedia. Perkiraan kedua jumlah jumlah kesalahan yang melekat pada nilai tes tertentu dan seseorang benar skor, dua statistik diperlukan: (1) suatu keandalan koefisien tertentu untuk menyamaratakan, dan (2) deviasi standar tes.

KOEFISIEN KEANDALAN

Simbol yang digunakan untuk menunjukkan keandalan coefficient adalah r dengan dua identik subskrip (misalnya, r_{xx} atau r_{aa}). Koefisien reabilitasi yang umumnya didefinisikan sebagai kuadrat diperoleh korelasi antara skor salah dan skor benar pada ukuran (r_{xt}^2). Ternyata, kuantitas ini identik dengan rasio varians dari skor benar varians dari skor yang diperoleh untuk suatu distribusi.

Karenanya, koefisien keandalan menunjukkan proporsionalitas variabilitas dalam suatu seperangkat nilai yang benar mencerminkan perbedaan di antara individu. Dalam kasus khusus di mana dua bentuk paralel tes ada, produk Pearson korelasi momen koefisien antara skor dari dua bentuk yang sama dengan keandalan koefisien untuk kedua bentuk. Hubungan ini diringkas dalam persamaan 7.1, dimana x dan x' adalah langkah-langkah paralel, dan S^2 adalah, kepastian varians.

$$r_{xx'} = r_{xt}^2 = \frac{S_{\text{angka benar}}^2}{S^2} = r_{xx'}$$

Sosroinei tampaknya Jika ada kesalahan relatif kecil, rasio varians angka yang benar untuk diperoleh varians skor mendekati indeks keandalan 1,00 (Quewct keandalan); jika ada adalah jumlah yang relatif besar kesalahan, rasio angka benar varians untuk obvariens skor tained-pendekatan ,00 (mm] unreliability). "Jadi, tes dengan koefisien reliabilitas ,90 memiliki kesalahan relatif kurang pengukuran dan lebih diandalkan daripada tes dengan keandala koefisien dari ,50.

Ada berbagai metode dalam memperkirakan sebuah koefisien keandalan yang bergantung pada satu generalisasi apa yang ingin dibuat. Test penulis harus selalu melaporkan sejauh mana orang bisa menyamaratakan ke berbagai kali dan tingkat yang satu dapat menggeneralisasi sampel yang berbeda pertanyaan atau item. Jika tes adalah sulit untuk mencetak, tes penulis juga harus melaporkan sejauh mana satu dapat menggeneralisasi skor yang berbeda.

Generalisasi untuk Berbeda Waktu

Test-retest keandalan adalah indeks drai kestabilan. Pendidik tertarik dalam banyak sifat dan karakteristik manusia itu, secara teoritis, berubah sangat sedikit di atas waktu. Sebagai contoh, anak-anak didiagnosis buta warna pada umur lima tahun diharapkan akan didiagnosa sebagai buta warna setiap saat dalam kehidupan mereka. Kekosongan warna adalah mewarisi sifat yang tidak dapat diperbaiki. Akibatnya, seharusnya sifat sempurna stabil. Ketika tes identitas seorang anak sebagai buta warna pada satu kesempatan dan tidak buta warna pada kesempatan berikutnya, tes ini tidak bisa diandalkan. Ciri lain kurang stabil daripada visi warna selama jangka waktu yang panjang, mereka adalah

perkembangan. Tinggi seseorang akan meningkat dari lahir sampai dewasa. Kenaikan relatif lambat dan mudah ditebak. Akibatnya, ukuran ment dengan penguasa yang andal harus menunjukkan perubahan kecil dalam ketinggian lebih dari satu bulan. Perubahan radikal tinggi (terutama berkurang) lebih pendek periode waktu yang akan menyebabkan kita untuk mempertanyakan keandalan pengukuran perangkat. Kebanyakan pendidikan dan karakteristik psikologis dikonseptualisasikan sebanyak ketinggian. Sebagai contoh, kita berharap prestasi membaca untuk meningkatkan dengan panjang sekolah tetapi menjadi relatif stabil selama periode waktu yang singkat, seperti dua minggu. Perangkat yang digunakan untuk menilai sifat dan karakteristik harus menghasilkan konsisten dan stabil suhiciently hasil jika hasil tersebut telah praktis berarti untuk membuat keputusan pendidikan.

Prosedur untuk mendapatkan koefisien stabilitas cukup sederhana. Besar jumlah siswa yang diuji. Tak lama kemudian (lebih dua minggu, tapi selang waktu dapat bervariasi dari satu hari ke beberapa bulan) mereka diuji ulang dengan perangkat yang sama. Para siswa skor dari kedua pemerintahan yang kemudian berkorelasi. koefisien korelasi yang diperoleh adalah rmbiligv caejicient Perkiraan jumlah kesalahan yang berasal dari kestabilan koefisien cenderung menjadi indated. Setiap perubahan pada siswa berkaitan dengan nilai benar pematangan akan ditambahkan ke varians kesalahan kecuali setiap mahasiswa dalam sampel perubahan dalam cara yang sama, Jadi, jika ada "pematangan semburan" antara dua tes administrasi hanya beberapa siswa, perubahan dalam skor yang benar adalah dimasukkan ke dalam istilah kesalahan. Demikian pula, jika beberapa siswa tidak dapat menjawab beberapa pertanyaan mengenai administrasi lirst ujian tetapi belajar jawaban oleh pemerintahan kedua, pembelajaran (perubahan dalam skor yang benar) diartikan sebagai kesalahan. Pengalaman mengambil ujian sekali mungkin juga membuat menjawab pertanyaan yang sama untuk kedua kalinya lebih mudah; tes pertama mungkin peka, siswa untuk administrasi kedua tes. Umumnya, semakin dekat bersama dalam waktu tes dan tes ulang adalah, semakin tinggi keandalan adalah, karena dalam jangka waktu yang lebih singkat ada kemungkinan kurang benar skor berubah.

Generalisasi untuk Sampel Item yang Berbeda

Ada dua pendekatan utama untuk memperkirakan sejauh mana kita dapat menggeneralisasi sampel berbagai item. Pendekatan pertama mengharuskan penulis mengembangkan dua (atau lebih) tes serupa, yang disebut pertanian alternatif yang kedua pendekatan tidak. Alternatif bentuk tes adalah detined

sebagai dua tes yang mengukur sifat yang sama atau keterampilan pada tingkat yang sama dan standar pada populasi yang sama. Alternatif bentuk tes setara pada dasarnya, kadang-kadang, pada kenyataannya, mereka disebut setara punts. Mari kita melihat sebuah contoh nonpsychometric. Lokal di berbagai toko counter, di mana beberapa 12 inci dijual penguasa, setiap penguasa dianggap sebagai setara (atau bentuk altemate) penguasa lain. Jika satu membeli penggaris merah dan hijau mler dan diukur beberapa objek dengan kedua, orang akan mengharapkan korelasi yang tinggi antara pengukuran dan hijau merah pengukuran. Contoh ini analog dengan bentuk alternatif kehandalan.

Ada satu perbedaan penting, namun. Alternatif fonns tes tidak berisi item yang sama. Tetap saja, sementara item yang berbeda, sarana dan varians untuk kedua tes diasumsikan (atau seharusnya) sama. Dalam tidak adanya kesalahan pengukuran, setiap topik akan diharapkan untuk mendapatkan sarne skor pada kedua bentuk.

TABEL 7,1 Hypothetical Kinerja Dua puluh Anak-anak pada Test Item Sepuluh_

Total	ram	Evens	Mustahil	Anak	1	2	3	4	5	6	7	8	9	10	Test	Benar	Benar
1	+	-1	-	+	m	+	-	-	-	+	-	5	1	4			
2	+	+	+	+	-	+	+	+	7	+	8	5	3				
3	+	+	-	+	+	+	+	-	+	+	B	4	4				
4	+	+	+	+	+	+	+	+	-	+	9	5	4				
5	+	+	+	+	+	+	+	+	+	+	9	4	5				
6	+	+	-	+	-	+	+	+	+	+	8	5	3				
7	+	+	+	+	+	-	+	-	+	+	8	3	5				
S	+	+	+	7	+	+	+	+	+	+	9	4	5				
9	+	+	+	+	+	+	-	•	+	+	+	9	5	4			
10	+	+	+	+	+	-	+	+	+	+	+	9	4	5			
11	+	+	+	+	+	-	+	-	-	-	6	2	4				
12	+	+	-	+	+	+	+	+	+	+	9	5	4				
13	+	+	+	7	-	+	-	+	-	-	5	3	2				
14	+	+	+	+	+	+	+	-	+	+	9	4	5				
15	+	+	•	+	+	-	----				4	2	2				
16	-1	»+	+	+	+	+	+	+	+	+	10	5	5				
17	+	7	+	V	-	-	-	-	-	-	2	0	2				
18	+	7	+	+	+	+	+	+	+	+	9	4	5				
19	+	+	+	+	-	+	+	+	+	+	9	5	4				
20	+	-	#	-	#	+	-	+	-	-	3	2	1				

Untuk memperkirakan koefisien keandalan oleh dua bentuk alternatif (A dan B), sebuah sampel besar siswa diuji dengan kedua bentuk. Setengah subyek menerima formulir A, maka form B setengah lainnya menipu form H, kemudian bentuk A. Skor dari kedua bentuk tersebut berkorelasi. Koefisien korelasi adalah koefisien kehandalan. Perkiraan didasarkan pada alternatif kehandalan bentuk tunduk pada salah satu kendala yang sama adalah stabilitas koefisien: semakin banyak waktu antara administrasi trationof dua (atau lebih) bentuk, semakin besar kemungkinan ofehghe sejati skor. Tidak seperti koefisien stabilitas, keandalan alternatif-bentuk perkiraan tidak dikenakan sensitisasi etfect sejak mata pelajaran tidak diuji dengan yang sama item dua kali.

Pendekatan kedua tidak mengharuskan para penulis lebih dari mengembangkan salah satu bentuk ujian. Metode ini memperkirakan keandalan suatu tes yang disebut ` konsistensi adalah sedikit berbeda. Misalkan kita ingin menggunakan metode kedua ini memperkirakan tes's reliability. Untuk melakukannya kita tes constmet berisi sepuluh item dan melaksanakan ujian dua puluh anak-anak. Para rlesillts tes hipotetis ini disajikan pada Tabel 7.1. Jika tes individu sepuluh item al.1 mengukur sifat atau karakteristik yang sama, kita dapat membagi tes menjadi dua tive-item tes, masing-masing berukuran sifat yang sama atau karakteristik. Dengan demikian, 1 tes ini diberikan, kita dapat membuat dua bentuk alternatif pengujian, masing-masing berisi satu-setengah dari jumlah tesitem, atau lima item. Kita kemudian dapat menghubungkan dua set nilai dan memperoleh perkiraan keandalan dari masing-masing dari dua bagian dengan cara yang sama kita memperkirakan akan keandalan dari dua alternatif bentuk ujian. Prosedur ini untuk memperkirakan tes keandalan yang disebut split-jerami reliabilmz perkiraan.

Harus jelas bahwa ada banyak cara untuk membagi tes menjadi dua sama-tes panjang. Sepuluh item pengujian dalam Tabel 7.1 dapat dibagi menjadi 126 tive pasang berbeda-item tes) Jika sepuluh item dalam tes penuh kita diatur dalam rangka peningkatan dilliculty, kedua bagian harus berisi item dari ofthe tes awal (yaitu, mudah item) dan item dari akhir tes ofthe (lebih keras item). Ada banyak cara untuk membagi pengujian seperti (misalnya 1, 4, 5, S, 9, dan 2, 3, 6, 7, 10). Cara yang paling umum untuk membagi sebuah tes adalah dengan ganjil dan bahkan-ntunbered item (lihat kolom berlabel "Evens Benar "dan" Mustahil Benar "dalam Tabel 7.1).

Sementara aneh-bahkan divisi dan kemudian korelasi dari dua bagian dari suatu tes adalah metode umum untuk memperkirakan tes konsistensi internal-reliability, mereka tidak selalu menawarkan metode terbaik. Pada kenyataannya, tergantung pada bagaimana tes ini dibagi menjadi dua bagian, perkiraan keandalan

akan bervariasi, yang lebih metode digeneralisasi memperkirakan konsistensi internal yang telah dikembangkan oleh Cronbach (1951) dan disebut koefisien internal. Koefisien alfa adalah rata-rata split-setengah korelasi berdasarkan semua kemungkinan pembagian tes menjadi dua bagian. Di praktek tidak perlu untuk menghitung semua kemungkinan koefisien korelasi; koefisien alfa dapat dihitung dari varians item tes individu dan varians dari total skor tes seperti yang ditunjukkan dalam persamaan 7,2 di mana k adalah jumlah item dalam ujian.

$$r_{aa} = 1 - \frac{1}{k} \sum r_{kk} \quad (7.2)$$

Koefisien alfa dapat digunakan ketika item tes lulus-gagal mencetak atau ketika lebih dari satu titik kredit diberikan untuk jawaban yang tepat. Sebelumnya, metode lebih ketat memperkirakan keandalan suatu tes, metode yang didasarkan pada rata-rata korelasi antara semua kemungkinan terbelah bagian, dikembangkan oleh Kuder dan Richardson. Prosedur ini disebut KR-20 dan koefisien alfa untuk mencetak tes dichotomously item (yaitu, barang yang dapat mencetak gol hanya benar ~ atau salah). Persamaan 7,2 dapat digunakan dengan data dikotomis, namun dalam hal ini misal perkiraan hasil reliabilitas biasanya disebut KR-20 perkiraan.

$$r_{aa} = \frac{1}{k} \sum r_{kk} \quad (7.2)$$

Ada dua pertimbangan utama dalam penggunaan internal konsistensi estimasi pasangan. Pertama, metode ini tidak boleh digunakan untuk tes atau tes berjangka waktu yang tidak diselesaikan oleh semua orang yang sedang diuji. Kedua, tidak memberikan perkiraan stabilitas dari waktu ke waktu. $LZS = \frac{1}{5!} \sum V$ J. Kadang-kadang penulis akan memperkirakan tes KR-20 dengan formula yang disebut KR-21. I-lewever, ini sebagai. biasanya nat rheneul yang diinginkan.

Table 7.2 pendapat perubahan bentuk gambar pada golongan emosi yang mengganggu anak

Nomor anak	Ny. Hawthorne	Torrance
1	normal	normal
2	berubah	berubah
3	berubah	normal
4	normal	normal
5	normal	normal
6	berubah	berubah

7	berubah	berubah
8	berubah	normal
9	normal	normal
10	normal	berubah
11	berubah	berubah
12	normal	normal
13	normal	normal
14	normal	normal
15	berubah	berubah
16	normal	berubah
17	normal	normal
18	berubah	berubah
19	normal	berubah
20	normal	berubah

PENYAMARATAAN UNTUK PERBEDAAN ANGKA

Ada dua cara yang sangat berbeda dalam penilaian tingkat yang mana kita dapat menyamakan perbedaan angka. Cara yang pertama serupa dengan cara penilaian yang dapat disamaratakan itu kita tepat melakukan diskusi. Kedua penguji angka menentukan ujian secara bebas. Angka diperoleh dari tiap-tiap penguji untuk menentukan lalu dihubungkan. Hasil hubungan koefisien adalah koefisien yang dapat dipercaya sebagai angka. Sebagai contoh, andai kata seorang psikolog (Ny. Hawthorne) berkepentingan dalam penyimpangan bentuk gambar pada emosional yang mengganggu anak sekolah. Selanjutnya andai kata dia ditentukan untuk menilai penyimpangan pada penilaian gambar sosok manusia seperti anak. Sama dengan jelas patokan untuk mengangkat gambar yang menyimpang, angka pada gambar sosok manusia sangat susah. Apakah yang lain, sama-sama terlatih, penguji Torrance mencapai kesimpulan yang sama dengan Ny. Hawthorne ? Dapatkah pendapat dan angka Ny. Hawthorne disamakan dengan penguji lain ? untuk mengukur tingkat yang mana jenis pada

penyamarataan ini tepat, dua penguji dapat menilai gambar sosok manusia dibuat oleh golongan pada emosional yang mengganggu anak-anak. Ditunjukkan pada table 7.2 disana akan menjadi dua penilaian penyimpangan bentuk gambar pada setiap gambar, dan dua angka dapat dihubungkan. Hasil hubungan koefisien ($\phi =$) akan menjadi sebuah penilaian interscore yang dapat di percaya atau mufakat (persetujuan).

Metode yang kedua pada penilaian penyamarataan untuk angka yang berbeda adalah merata pada penilaian tingkah laku. Sebagai pengganti menghubungkan dua angka penilaian, satu dapat memperhitungkan persentasi pada persetujuan. Petunjuk sederhana pada persetujuan diperoleh dari pemisahan angka yang disetujui dengan angka yang tidak disetujui dan mengalikan hasil bagi dengan 100. untuk data pada table 7.2, persentasi pada persetujuan adalah 70% [yaitu, $(14/20) (100)$]. Cara yang lain dalam menghitung persentasi persetujuan adalah menghitung persentasi persetujuan selama kejadian pada sasaran tingkah laku (lihat persamaan 7.3). pada contoh ini, sejak Ny. Hawthorne dihubungkan ke dalam kejadian penyimpangan pada bentuk gambar, Ny Hawthorne boleh merasa lebih baik untkhanya melihat bagaimana baiknya dua angka setuju pada kejadian. Delapan bukan kejadian dimana Ny. Hawthorne dan Torrance setuju tidak memperhatikan dan diabaikan. Data pada table 7.2 dapat diringkas pada table 7.3. persentasi persetujuan untuk kejadian adalah 50% [adalah, $(100)(6)/(20-8)$].

% persetujuan kejadian =

$$\frac{(100)(\text{angka persetujuan pada kejadian})}{\text{Angka pada penelitian} - \text{angka persetujuan pada bukan kejadian}} \quad (7.3)$$

METODE APA YANG HARUS DIGUNAKAN ?

Pertimbangan pertama adalah pada jenis penyamarataan yang ingin di buat. Satu harus memilih metode yang berjalan dengan penyamarataan. Sebagai contoh, jika dimana yang satu berkepentingan menyamaratakan tentang keseimbangan pada angka atau pengamatan, metode yang tepat akan menjadi hubungan test-retest. itu akan menjadi tidak tepat untuk menggunakan persetujuan interscore pada penilaian perpanjangan yang mana satu dapat menyamakan untuk perbedaan waktu. Pertimbangan tambahan pada memilih metode yang digunakan memasukan berikut :

1.ketika penilaian seimbang, perjanjiannya adalah untuk retest setelah dua minggu. Disana tidak ada yang khusus dari dua minggu. (jika semua test

pengarang menggunakan jarak yang sama, itu akan menjadi mudah untuk membandingkan relative keseimbangan pada test.

2.ketika penilaian perpanjangan yang mana untuk kita dapat menyamaratakan untuk hal test yang serupa, kita menganut Nunnally's (1967, p.217) hirarki untuk penilaian yang dapat dipercaya. Pilihan pertama adalah menggunakan bentuk pengganti yang banyak dipercaya dengan jarak dua minggu. (sekali lagi, di sini tidak ada yang khusus tentang dua minggu ; ini hanya adapt). Jika bentuk pengganti tidak tersedia, membagi test ke dalam dua bagian yang sama dan menjalankan dua bagian dengan jarak dua minggu, perbaikan hubungan oleh spearman- brown memberikan rumus pada persamaan 7.4. ketika bentuk pengganti tidak tersedia dan pokok persoalan tidak dapat di test lebih dari satu kali, gunakan koefisien alfa.

3.ketika penilaian perpanjangan yang mana kita dapat menyamaratakan diantara perbedaan angka, kita lebih suka menghitung korelasi koefisien untuk persentasi pada persetujuan. Korelasi koefisien menunjang kelangsungan hubungan penunjuk yang lain yang dapat dipercaya dan penggunaan yang lain pada koefisien yang dapat dipercaya ; persentasi tidak disetujui. Kita juga menyadari bahwa pada prakteknya sekarang adalah laporan persentasi persetujuan dan tidak mengusahakan dengan menggunakan koefisien yang dapat dipercaya untuk menyamaratakan angka yang lain.

FAKTOR YANG MEMPENGARUHI RELIABILITY

Beberapa factor yang mempengaruhi test reliability dapat membumbung atau mengempis penilaian reliability.

LAMANYA TEST

Peraturan pada umumnya, banyak hal yang sama pada test, test lebih dapat dipercaya. Demikian, lamanya test menjaga lebih dapat dipercaya dari pada sebentarnya atau pendeknya test. Kenyataan ini penting terutama pada ketetapan internal penilaian yang dapat dipercaya. Karena pada jenis penilaian ini angka pada semua test dikurangi dengan 50%. Ketetapan internal penilaian pada penililaian yang benar-benar dapat dipercaya kepercayaan pada separuh test. Oleh karena itu, penilaian seperti itu sering dibenarkan dengan rumus yang dikembangkan oleh spearman dan brown. Ditunjukkan pada persamaan 7.4., kepercayaan total test adalah sama untuk dua kali kepercayaan pada penilaian dengan ketetapan internal dibagi dengan jumlah satu ditambah penililaian yang dapat dipercaya.

$$R_{xx} = 2r(1/2) (1/2) = 1.60$$

$$1+r(1/2)(1/2) \quad 1.80$$

Sebagai contoh, jika koefisien alfa dimana dihitung pada test dan ditemukan menjadi 80, dinilai dibenarkan dapat dipercaya akan menjadi .89,

$$.89 = (2)(.80) = 1.60$$

$$1+ .80 \quad 1.80$$

Hubungan persoalan adalah angka pada hal yang berhasil pada tiap-tiap pengambil test. Penguji pada umumnya lebih dapat dipercaya pada jarak pertengahan pada angka (sebagai contoh, $\pm 1.5S$). agar test menjadi berhasil pada penyaluran yang berbeda, mungkin akan menjadi cukup angka pada hal yang berbeda agar anak lebih cerdas seperti angka yang cukup merupakan hal yang mudah bagi anak yang terbelakang.

JARAK TEST-RETEST

Banyak orang membenarkan kecakapan dapat dan melakukan perubahan diantara dua administrasi pada test. Jumlah yang besar pada waktu diantara dua administrasi, mungkin lebih banyak kemungkinan angka yang benar akan berubah. Demikian, ketika menggunakan keseimbangan atau bentuk pengganti penilaian yang dipercaya, satu harus menutup upah perhatian pada jarak antara test. Umumnya, pendeknya jarak, tingginya penilaian yang dapat dipercaya.

PENYEMPITAN JARAK

Penyempitan jarak terjadi ketika contoh dengan perbedaan relative kecil digunakan untuk penilaian test yang dapat dipercaya. Ketika contoh dengan dikurangi jarak yang digunakan, test yang dipercaya akan menjadi sungguh-sungguh diremehkan karena koefisien yang dapat dipercaya langsung dihubungkan pada variability pada test. Perbedaan yang sangat besar pada test, penilaian yang dapat dipercaya juga besar.

Pada bagan 7.1 ditunjukkan, bentuk-bentuk pilihan pada hasil test sebuah hubungan positif yang kuat ketika seluruh jarak digunakan. Bagaimanapun, dalam beberapa jarak terbatas pada test, digambarkan dengan garis bentuk empat persegi panjang, hubungan mungkin sangat rendah. (walaupun kemungkinan untuk memperbaiki koefisien korelasi selama pembatasan pada jarak, pada umumnya ini tidak bijaksana untuk dilakukan juga.)

Masalah yang dihubungkan adalah perpanjangan pada jarak sungguh-sungguh terlalu tinggi menilai test kepercayaan. Bagan 7.2 mengandung bentuk pengganti korelasi yang pertama, ketiga dan kelima angka. Persebaran pada tiap-

tiap angka, dipertimbangkan terpisah, menunjukkan kepercayaan yang tidak baik. Bagaimanapun, test ejaan bertambahnya angka pada fungsi sekolah ; murid-murid mendapat angka yang besar mendapat nilai yang besar. Ketika penguji menyatukan angka-angka untuk beberapa nilai (dari beberapa usia), hubungan yang tidak baik mungkin disatukan untuk hasil yang palsu tinggi hubungannya.

TEKA-TEKI (PERKIRAAN)

Teka-teki adalah menjawab hal-hal yang serampangan. Tetap jika hasil perkiraan jawaban benar, memasukan kesalahan ke dalam angka test dan ke dalam taksiran kita pada angka.

PERBEDAAN DALAM PENGUJIAN KEADAAN

Banyaknya kesalahan bahwa perbedaan pada test keadaan memasukan ke dalam hasil pada test dapat mengubah peraturan. Murid- murid dapat salah membaca atau salah paham dengan petunjuk pada test, memperoleh soal separuh jalan selesai test, bagian yang hilang pada lembar jawaban, keretakan pada ujung pensilnya, atau memilih untuk menjaga tupai makan kacang-kacangan pada ambang jendela ruang kelas lebih baik dari pada pengambilan test. Semua seperti perbedaan keadaan memasukan jumlah kesalahan yang tak menentu pada test dan pada yang dilakukannya juga menurunkan kepercayaan.

STANDAR KESALAHAN PADA PENGUKURAN

Alasan yang paling utama untuk memperoleh koefisien reliable adalah menilai jumlah kesalahan biasanya dihubungkan dengan penyamarataan dari angka yang diperoleh dibawah kumpulan fakta pada keadaan. Standar kesalahan pada pengukuran (SEM) mengijinkan satu penilaian kesalahan ditemani dengan tiap-tiap jenis kesalahan. Satu dapat memperhitungkan standar kesalahan pada pengukuran pada angka-angka, waktu-waktu dan contoh-contoh hal. Bagaimanapun SEM biasanya dihitung hanya pada keseimbangan dan contoh-contoh hal.

Kita lebih dahulu membicarakan penyamarataan pertunjukan pada contoh hal untuk rata-rata semua contoh kemungkinan pada penarikan hal dari daerah kekuasaan. Penyediaan ini merupakan contoh yang tepat pada standar kesalahan pada perhitungan. Menganggap abjad mengakui abjad lagi. Disana banyak contoh pada sepuluh surat test dapat menjadi dikembangkan. Jika kita dibangun hanya 100 pad test dan test pada satu taman kanak-kanak, kita mungkin menemukan penyaluran angka pad ataman kanak-kanak kira-kira normal. Rata-rata pada penyaluran akan menjadi angka benar pada anak-anak. Penyaluran sekitar angka yang benar akan menjadi contoh hasil yang tidak sempurna pada surat ; beberapa

contoh surat akan melebihi penilaian kemampuan anak-anak, dan akan melebihi penilaian yang lain. Demikian, penyaluran akan menjadi hasil yang salah. Standar deviasi pada penyaluran standar deviasi kesalahan diakibatkan penarikan contoh dan disebut standar kesalahan pada pengukuran.

Ketika kita menguji murid, kita khusus menguji hanya sekali. Oleh karena itu kita tidak dapat menyamakan penyaluran serupa pada satu yang digambarkan bagan 7.3. maka dari itu kita tidak mengetahui pengambilan test angka yang benar atau dapat dipercaya pada ukuran kesalahan pada bentuk penyaluran sekitar angka yang benar. Bagaimanapun kita dapat menggunakan apa yang kita tau tentang kepercayaan test pada hal-hal dan standar deviasi untuk menilai apa yang akan menjadi penyaluran.

Persamaan 7.5 adalah rumus umum pada penemuan standar kesalahan pada pengukuran. Standar kesalahan pada pengukuran (SEM) sama dengan standar deviasi diperoleh angka (S) dikalikan dengan kuadrat akar pada satu dikurangi koefisien realibel (akar dari $1-r_{xx}$). Jenis kesatuan (IQ, nilai mentah, dan sebagainya) yang mana pada standar deviasi adalah tepat yang mana kesatuan pada SEM adalah tepat. Demikian, jika angka test dirubah menjadi angka T, standar deviasi adalah kumpulan angka T adalah 10 ; SEM juga pada kumpulan angka T. sama halnya, jika koefisien reliable didasarkan pada keseimbangan. Lalu SEM selama waktu test. Jika koefisien reliable didasarkan pada perbedaan angka-angka, lalu SEM adalah untuk pengujian.

Dari persamaan 7.5 jelas kelihatan pertambahan standar deviasi, SEM bertambah ; dan koefisien reliable berkurang, SEM bertambah. Pada bagian A tabel 7.4 standar deviasi sama (10) digunakan dengan perbedaan koefisien realibel. Koefisien realibel berkurang, SEM bertambah. Ketika koefisien realibel adalah .96, SEM adalah 2 ; ketika reliable adalah .64, SEM adalah 6. pada bagian B pada tabel 7.4, perbedaan standar deviasi digunakan dengan koefisien reliable yang sama ($r_{xx} = .91$). standar deviasi bertambah, SEM bertambah.

Karena kehadiran kesalahan pada pengukuran, disana selalu beberapa ketidakpastian tentang kebenaran angka tersendiri. Standar kesalahan pada pengukuran menyediakan informasi tentang kepastian atau kepercayaan dengan yang mana angka test dapat diartikan. Ketika SEM relative besar, ketidakpastian besar ; kita tidak dapat menjadi sangat yakin pada angka seseorang. Ketika SEM relative kecil, ketidakpastian kecil ; kita dapat menjadi lebih pasti dengan angka.

PENILAIAN KEBENARAN ANGKA

Sayang, kita tidak dapat mengetahui pokok kebenaran angka. Selain itu, diperoleh angka pada test tidak lebih baik penilaian kebenaran angka. Disebutkan pada pembicaraan sebelumnya, kebenaran angka dan kesalahan tidak ada hubungannya. Bagaimanapun, diperoleh angka dan kesalahan tidak berhubungan. Angka di atas pada rata-rata test lebih “ untung “ kesalahan (kesalahan kenaikan diperoleh angka di atas kebenaran angka), sementara angka bawah rata-rata lebih “ tidak untung “ kesalahan (kesalahan dibawah diperoleh angka rendah kebenaran angka). Cara yang mudah untuk memahami pengaruh ini adalah memikirkan test pada yang mana perkiraan murid pada setengah hal test. Jika semua perkiraan benar, murid menjadi lebih beruntung dan memperoleh angka yang tinggi.

TABEL 7,5 Benar Perkiraan Skor untuk Berbagai Memperoleh Skor pada Pengujian dengan Berbeda Koefisien Reliabilitas Selisih Antara Uji Reliabilitas Memperoleh Memperoleh Perkiraan Nilai

Test Rata-rata (X)	Koefisien Reabilitas (r xx)	Skor Obstain (X)	Skor Benar (X)	Perkiraan Benar dan Nilai Skor
--------------------	-----------------------------	------------------	----------------	--------------------------------

100	.90	90	91,0	1,0
100	.90	75	77,5	2,5
100	.90	50	55,0	5,0
100	.70	90	93,0	3,0
100	.70	75	82,5	7,5
100	.70	50	65,0	1
100	.50	90	95,0	5,0
100	.50	75	87,5	12,5
100	.50	50	75,0	25,0

Namun, jika semua dugaan-dugaan tidak benar, mahasiswa telah beruntung dan menghasilkan kelas rendah. Dengan demikian, diperoleh skor di atas atau di bawah rata-rata sering kali lebih berbeda-beda dari nilai benar. Seperti dapat dilihat dari Gambar 7.4, yang kurang dapat diandalkan tes, semakin besar

perbedaan antara nilai yang diperoleh dan benar skor. Nunnally (1967, hal 220) telah memberikan suatu persamaan (persamaan 7.6) untuk determining perkiraan nilai sejati (X'). 'Tia memperkirakan nilai benar berarti sama dengan tes ditambah produk dari koefisien keandalan dan perbedaan antara diperoleh skor dan kelompok berarti.

$$X = X + (r_{.}) < X - X \text{ (76)}$$

Tertentu yang berarti bahwa satu menggunakan adalah subjek dari beberapa kontroversi. Kami percaya pilihan mean mean dari grup demografis yang paling mewakili anak tertentu. Jadi, jika anak Asia dan tinggal di kelas bawah wilayah perkotaan, yang paling sesuai akan berarti mean Asia usia sama anak-anak dari latar belakang sosial-ekonomi rendah yang tinggal di perkotaan. Dalam ketiadaan sarana untuk anak-anak tertentu tertentu latar belakang, satu dipaksa untuk menggunakan keseluruhan mean untuk usia anak. Sebagaimana telah terjadi di seluruh bab ini, pilihan reliabilitas coefficient tergantung pada jenis generalisasi seseorang berkeinginan untuk membuat. Perbedaan antara nilai yang diperoleh dan estimasi nilai benar adalah fungsi dari kedua THC ofthe kehandalan diperoleh skor dan perbedaan antara andthe skor yang diperoleh berarti. Dalam Tabel 7.5, kasus umum digambarkan di mana rata-rata di setiap contoh adalah 100; yang diperoleh skor 90, 75, dan 50. Itu kehandalan coefficients adalah ,90, ,70, dan .50. Bila nilai yang diperoleh adalah 90 dan diperkirakan keandalan adalah ,90, perkiraan nilai sejati adalah 91 [$91 = 100 + (.90) (90 - 100)$]. Namun, ketika skor yang diperoleh adalah 50 dan kehandalan coefficient adalah ,90, perkiraan nilai sejati adalah 55 [$100 + (.90) (50 - 100)$]. Bahkan ketika kehandalan coefficient adalah konstan, semakin jauh sebuah skor yang diperoleh dari berarti, semakin besar akan perbedaan antara nilai yang diperoleh dan perkiraan skor benar. Bila nilai yang diperoleh adalah 75 dan kehandalan coefficient adalah ,90, yang diperkirakan benar skor 77,5 [$100 + (.90) (75 - 100)$]. Namun, ketika koefisien keandalan tetes untuk ,50 dan skor yang diperoleh tidak berubah, skor benar perkiraan meningkat menjadi 87 [$100 - 4 - (.50) (75 - 100)$]. Ketika diperoleh skor tes di bawah rata-rata dan koefisien reliabilitas kurang dari 1,00, perkiraan nilai sejati selalu lebih tinggi daripada yang diperoleh skor.

Sebaliknya, ketika skor yang diperoleh berada di atas rata-rata dan uji koefisien reliabilitas kurang dari 1,00, perkiraan nilai sejati selalu kurang daripada skor yang diperoleh. Perhatikan bahwa persamaan tidak memberikan ketakutan yang benar, hanya perkiraan benar ketakutan,

KEPERCAYAAN INTERVAL

Meskipun kita tidak pernah tahu seseorang benar skor, pengukuran bukan aktivitas putus asa. Kita dapat memperkirakan nilai yang benar, dan kita dapat memperkirakan deviasi standar kesalahan pengukuran mengenai skor benar. Dengan dua bit informasi, kita dapat membuat berbagai di mana kita tahu probabilitas yang tepat termasuk orang benar skor. Kisaran ini disebut sebagai confidence selang interval kepercayaan 50 persen adalah serangkaian nilai-nilai dalam skor yang sejati akan ditemukan 50 persen dari waktu. Tentu saja, 50 persen waktu skor yang benar akan berada di luar interval. Rentang yang lebih besar ~ yang lebih besar confidence interval-bisa membuat kita merasa lebih yakin bahwa kami telah memasukkan skor yang benar dalam jangkauan. Tetapi tidak mungkin untuk membangun suatu interval di mana nilai sejati akan selalu terkandung. Namun, jika kita membangun 95 persen atau 99 persen interval keyakinan, maka kemungkinan hanya 5 persen dan satu persen, masing-masing, bahwa skor yang benar akan jatuh di luar keyakinan interval. Membentuk Confidence untuk True Interval Skor Karakteristik kurva normal telah dibahas. Kami akan menerapkan hubungan antara z-skor dan area di bawah kurva normal distribusi normal kesalahan sekitar skor sejati. Kita dapat menggunakan persamaan 7,6 untuk memperkirakan mean dari distribusi (skor yang benar) dan persamaan 7,5 memperkirakan deviasi standar dari distribusi (standard error pengukuran). Dengan kedua perkiraan, kita dapat membangun kepercayaan interval untuk skor sejati. Sejak 68 persen dari semua elemen dalam distribusi normal jatuh dalam satu standar deviasi dari mean, ada 68 persen kesempatan.

TABEL 7,6 Sering Digunakan z-Scores, Extreme Daerah, dan Area Termasuk Antara + dan • z-Skor

z-Score	Extreme Area	Antara Area + dan -
0,67	25,0%	50%
1,00	16,0%	68%
1,64	5,0%	90%
1,96	2,5%	95%
2,33	1,0%	98%
2,57	.5%	99%

Bahwa nilai yang benar dalam waktu satu SEM dari perkiraan skor sejati. Kita dapat membangun interval dengan hampir semua kecuali tingkat kepercayaan 100 persen keyakinan. Tabel 7.6 berisi daerah ekstrim untuk 1-skor paling commonly digunakan dalam interval keyakinan constructing. Rumus umum untuk interval keyakinan diberikan dalam persamaan 7.7. Batas bawah dari keyakinan Interval sama dengan perkiraan nilainya kurang benar produk dari z-skor sebagai - sociated dengan tingkat kepercayaan dan kesalahan standar pengukuran. Batas atas interval kepercayaan adalah benar perkiraan nilai plus produk dari z-skor dan SEM. Batas bawah c.i. = $\bar{X} - (z \cdot \text{skor}) (\text{SEM})$ Batas atas c.i. ; $\bar{X} + (z \cdot \text{skor}) (\text{SEM})$ (7,7) Untuk membangun simetris interval kepercayaan untuk nilai sejati, sederhana prosedur diikuti.

1. Pilih tingkat keyakinan, misalnya, 95 persen.
2. Carilah z-skor yang terkait dengan tingkat kepercayaan. (Sebagai contoh, sebuah interval keyakinan 95 persen adalah antara z-skor dari - 1,96 dan + 1,96.)
3. Kalikan setiap z-skor yang terkait dengan interval keyakinan (misalnya, 1,96 untuk 95 persen keyakinan) oleh SEM.
4. Carilah perkiraan nilai sejati. > V 5. Ambil produk dari z-skor dan SEM, dan keduanya menambahkannya ke dan kurangi dari perkiraan skor sejati. Sebagai contoh, asumsikan bahwa diperkirakan persorrs skor sejati adalah 75 dan bahwa SEM adalah 5. Lebih lanjut berasumsi bahwa anda ingin menjadi 68 persen yakin membangun interval yang akan berisi nilai yang benar. Enam puluh delapan persen dari waktu, skor benar akan terkandung dalam interval 70 hingga 80 [$75 - (1) (5)$ untuk $75 + (1) (5)$]; ada 16 persen kemungkinan bahwa nilai yang benar

kurang dari 70 dan 16 persen kemungkinan bahwa nilai sebenarnya lebih besar dari SO. Jika Anda tidak mau menjadi salah 32 persen dari waktu, Anda harus meningkatkan lebar confidence interval. Jadi, dengan nilai benar yang sama (75) dan SEM (5), jika Anda ingin 95 persen keyakinan, ukuran interval perlu ditingkatkan; itu akan harus berkisar 65-85 [$75 - (1.96) (5)$ sampai $75 + (1.96) (5)$]. Sembilan puluh lima persen waktu ofthe nilai sebenarnya akan terkandung di dalam selang; sana adalah 2,5 persen kemungkinan bahwa nilai yang benar kurang dari 65, dan ada 2,5 persen kemungkinan bahwa lebih besar dari 85.

PERBEDAAN SCORES

Dalam banyak diterapkan pengaturan, kita tertarik pada perbedaan (perbedaan) menjadi - tween dua nilai. Sebagai contoh, kita mungkin ingin tahu apakah prestasi anak usia sepadan (sama) untuk usia mentalnya. Dalam banyak definisi educa - gangguan nasional (misalnya, ketidakmampuan belajar) suatu "signif • icant" perbedaan ditetapkan. Gangguan lain (misalnya, keterbelakangan mental) signifikan perbedaan tidak diharapkan. Ada beberapa pertimbangan utama dalam mengevaluasi "makna" dari perbedaan (Salvia & Bagus, 1982); yang keandalan dari perbedaan, kelangkaan perbedaan, dan psikologis V kebermaknaan ofthe diderence. Di sini, penekanannya adalah pada keandalan ofthe perbedaan. Yang penting untuk diingat adalah bahwa, biasanya, perbedaan nilai kurang dapat diandalkan dibandingkan dengan skor yang didasarkan perbedaan. Keandalan dari perbedaan antara dua nilai (A dan B) adalah fungsi dari empat hal: (1) keandalan tes A, (2) keandalan tes B, (3) korelasi antara tes A dan B, dan (4) perbedaan-perbedaan dalam kelompok norma. Ada adalah beberapa pendekatan untuk mengevaluasi keandalan dari suatu perbedaan. Dua meth - ods akan sangat berguna, tapi bertumpu pada asumsi yang berbeda dan menggabungkan data dalam berbagai cara (yaitu, gunakan rumus yang berbeda). Salah satu metode menggunakan model regresi dan awalnya digambarkan oleh Thorn - tanggul (1963). Dalam model ini, suatu skor diduga menyebabkan kedua skor. Sebagai contoh, kecerdasan diyakini menyebabkan prestasi. Oleh karena itu, intelijen diidentifikasi sebagai independen (atau prediksi) variabel, dan mencapai-2

ment diidentifikasi sebagai dependen (atau memprediksi) variabel. Ketika diprediksikan; skor (misalnya, nilai prestasi yang diprediksi) berbeda dari mencapai-` ment skor yang sebenarnya diperoleh, ada defisit. Keandalan dari pra - Perbedaan dicted diberikan oleh persamaan .

Keandalan dari prediksi berbeda - ence (D) adalah sama dengan keandalan dari variable dependen (r_{bb} ,) ditambah dengan • keandalan produk dari variabel independen (r_{aa}) dan Q korelasi antara variabel independen dan variabel dependen _ (r_{ab}) kurang dua kali kuadrat korelasi independen dan bergantung 4 variabel ($r_{aa} r^2_{ab}$). Kombinasi ini dibagi dengan satu minus kuadrat, _ korelasi antara independen dan variabel dependen ($1 - r_{ab}$). Standar deviasi perbedaan diperkirakan (S), juga disebut standard error dari estimasi, diberikan dalam persamaan 7.9. Deviasi standar prediksi perbedaan adalah samadengan deviasi standar variabel dependen (S_{bb}) dikalikan dengan akar kuadrat dari satu minus kuadrat korelasi antara independen dan variabel dependen ($\sqrt{1 - r^2_{ab}}$)

$$D = \frac{r_{aa} + (r_{bb}) - 2r_{ab}}{1 - r_{ab}} \quad 7.8.$$

$$S = s_{bb} \sqrt{1 - r^2_{ab}}$$

Metode kedua diusulkan oleh Saham dan Wardrop (1971). Dalam metode, satu variabel tidak diasumsikan menjadi penyebab yang lain, tidak variabel yang diidentifikasi sebagai variabel independen. Namun, metode ini tidak memerlukan langkah-langkah kedua berada dalam unit pengukuran yang sama (misalnya, T • skor atau IQ). Keandalan perbedaan dalam skor yang diperoleh diberikan dalam persamaan 7.10.

$$r_{dif} = \frac{1/2(r_{aa} + r_{bb}) - r_{ab}}{1 - r_{ab}}$$

Keandalan dari suatu perbedaan yang diperoleh sama dengan rata-rata ofthe keandalan dua tes [$1 / 2 (r_{aa} + r_{bb})$] korelasi antara kedua tes ($- r_{ab}$); perbedaan ini dibagi oleh satu minus korelasi antara kedua tes ($1 - r_{ab}$). Deviasi standar untuk perbedaan diperoleh diberikan dalam persamaan 7.11.

$$S_{\text{dif}} = \sqrt{S^2_a + S^2_b - 2r_{ab}S_aS_b} \quad (7,1)$$

Deviasi standar dari suatu perbedaan yang diperoleh sama dengan akar kuadrat varians sum of the of the tes A dan B ($S^2_a + S^2_b$) kurang dua kali lipat produk korelasi dari A dan B dikalikan dengan deviasi standar A dan B ($-2r_{ab}S_aS_b$). Keandalan perbedaan dan deviasi standar dikombinasikan dengan cara yang sama untuk perbedaan skor sebagai skor tunggal. Substitusi dalam persamaan 7.5, persamaan 7,12 dihasilkan.

$$SEM_{\text{dif}} = \sqrt{S^2_a + S^2_b - 2r_{ab}S_aS_b} \sqrt{1 - \frac{1}{2}(r_{aa} + r_{bb})}$$

Sebuah Error standar pengukuran perbedaan menggambarkan distribusi perbedaan antara nilai yang diperoleh. Mengevaluasi perbedaan nilai, yang simplest metode untuk menetapkan tingkat confidence (misalnya, 95%) dan menemukan haz-skor yang terkait dengan tingkat kepercayaan (1,96). Kami kemudian membagi perbedaan diperoleh oleh SEM perbedaan. Jika melebihi hasil bagi z-skor yang terkait dengan tingkat keyakinan yang dipilih (1,96), yang diperoleh difference dapat dipercaya. Ketika perbedaan diasumsikan dapat diandalkan pada tingkat tertentu kepercayaan, kita dapat memperkirakan perbedaan yang benar dengan cara yang sama seperti kita. memperkirakan nilai yang benar di uji. Secara umum, kita asumsikan bahwa kelompok berarti Perbedaannya adalah ,00. Dengan demikian, untuk memperkirakan formula sejati untuk difference mahasiswa tertentu untuk menyederhanakan persamaan 7.13. Perkiraan selisih sesungguhnya = (diperoleh perbedaan) (r,, (d, 0) (7.13)

DIINGINKAN STANDAR

Ini adalah ujian penting bagi penulis untuk menyajikan informasi yang memadai dalam tes manual bagi pengguna tes untuk menafsirkan hasil tes secara akurat. Hasil tes harus digeneralisasikan sebelum mereka berguna. Apakah suatu tes mengukur apa yang dimaksudkan untuk mengukur adalah sebuah pertanyaan tentang validitas, topik bab berikutnya. Akan tetapi, untuk tes yang akan berlaku (untuk mengukur apa yang para penulisnya mengklaim bahwa langkah-langkah),

harus dapat diandalkan. Meskipun bukan satu-satunya syarat yang harus dipenuhi, keandalan adalah syarat mutlak untuk validitas. Tidak ada tes dapat mengukur apa yang dimaksudkan ukuran kecuai `s dapat diandalkan. Tidak skor ditafsirkan kecuai `s dapat diandalkan. Sana kedepan, tes penulis dan penerbit harus menunjukkan data yang cukup untuk memungkinkan keandalan pengguna untuk menginterpretasikan hasil tes secara akurat. Indeks keandalan untuk tiap jenis dari skor (misalnya, nilai mentah, setara kelas, dan angka standar) harus dilaporkan. `ini harus dilaporkan untuk setiap usia dan kelas. Terlebih lagi, indeks ini harus disajikan dengan jelas dalam bentuk tabel di satu tempat. Tes penulis seharusnya tidak bermain petak umpet dengan keandalan data. Test penulis yang merekomendasikan komputasi dilference skor harus memberikan, jika memungkinkan. Keandalan dari perbedaan dan perbedaan SEM. Setelah tes pengguna memiliki akses ke data kehandalan, mereka harus menilai kecukupan ujian. Ketika nilai ujian dilaporkan, kami sangat merekomendasikan bahwa estimasi skor benar dan interval keyakinan 68 persen untuk itu benar juga skor dilaporkan. Seberapa tinggi harus suatu koefisien reliabilitas dapat sebelum dapat digunakan dalam pengaturan diterapkan? Tergantung gunakan untuk data pengujian yang dimasukkan. Sederhana jawabannya adalah dengan menggunakan tes yang paling dapat diandalkan tersedia. Namun, seperti respons menyesatkan, karena "terbaik" tes mungkin memiliki kehandalan coefficient terlalu rendah untuk setiap aplikasi (misalnya, ,12). Kami merekomendasikan bahwa dua standar keandalan digunakan dalam pengaturan diterapkan.

1. Kelompok Data Jika skor tes digunakan untuk keperluan administrasi dan dilaporkan untuk kelompok, yang keandalan ,60 sebaiknya menjadi minimum.
2. Individu Bendungan jika skor tes digunakan untuk membuat keputusan untuk satu siswa, yang jauh lebih tinggi standar kehandalan dituntut. Ketika penting educa – keputusan nasional, seperti pelacakan dan penempatan di dalam kelas khusus, adalah untuk dibuat untuk mahasiswa, standar minimum harus ,90. Ketika keputusan yang dibuat adalah keputusan skrining, seperti rekomendasi yang seorang anak menerima

penilaian lebih lanjut, masih ada kebutuhan untuk keandalan tinggi. Untuk perangkat skrining, kami merekomendasikan standar ,80.

RINGKASAN

Reliability mengacu pada kemampuan untuk menggeneralisasi dari sebuah contoh untuk sebuah domain, The domain yang biasanya kita ingin menggeneralisasi lain kali (stabilitas atau tes-tes ulang keandalan), skor lain (interrater atau interscorer keandalan), dan item lain (alternatif-bentuk atau kehandalan konsistensi internal). Reliabilitas co-efficient dapat berkisar dari ,00 (total kurangnya keandalan) sampai 1.00 (total keandalan); ,90 Dianjurkan sebagai standar minimum untuk tes digunakan untuk membuat penting keputusan pendidikan bagi anak-anak. Dalam diagnostik bekerja dengan anak-anak reli kemampuan coefficient memiliki tiga penggunaan utama; Hal ini memungkinkan pengguna (1) untuk memperkirakan tes relatif bebas dari kesalahan pengukuran, (2) untuk memperkirakan seorang individu skor benar subjek, dan (3) untuk mencari error standar pengukuran. Knowl - tepi kesalahan standar pengukuran dan perkiraan nilai rona memungkinkan pengguna tes untuk membangun kepercayaan interval untuk subjek skor benar. Diskusi perkiraan nilai benar, error standar pengukuran, dan interval keyakinan dapat diperpanjang untuk difference atau perbedaan skor. Keandalan dari skor difference dipengaruhi oleh keandalan dari tes dan oleh korelasi antara tes yang didasarkan perbedaan. Perbedaan sampel norma juga affect perbedaan nilai, tetapi efek ini uin - tidak dapat dievaluasi. Menyediakan dua tes tersebut berkorelasi, perbedaan skor kurang dapat diandalkan dibandingkan rata-rata tes reliabilities di mana perbedaan didasarkan. Ada beberapa faktor yang affect keandalan: metode yang digunakan untuk menghitung keandalan coefficient, panjang tes, tes-tes ulang interval, penyempitan jangkauan, menebak, dan variasi dalam situasi pengujian.

STUDI PERTANYAAN

1. Mengapa perlu bahwa tcst dapat diandalkan?

2. Tes tes A dan B memiliki identik berarti dan standar deviasi. Test A standar error pengukuran 4.8; tes B mempunyai standard error pengukuran 16.3. Pengujian yang lebih dapat diandalkan, dan mengapa?
3. Apakah pembatasan terbesar reliabilitas perkiraan berdasarkan tes-tes ulang korelasi?
4. Daftar dan menjelaskan faktor-faktor yang tive perkiraan affect keandalan tes.
5. Standard error pengukuran adalah deviasi standar apa? Mengilustrasikan jawaban Anda dengan gambar.

MASALAH

1. Mr Treacher mengelola tes kecerdasan ($X = 100$, $S = 16$, $r_{,,:} .75$) ke kelasnya. Lima anak memperoleh skor sebagai berikut: 68, 124, 84, 100, dan 148. Berapa perkiraan nilai yang benar bagi anak-anak ini?
2. Apa kesalahan standar pengukuran untuk tes kecerdasan masalah 1?
3. Apa saja batas-batas atas dan bawah dari sebuah keyakinan simetris interval 95 persen untuk anak pertama?
4. Apa saja batas-batas atas dan bawah dari sebuah keyakinan simetris Interval dari 50 persen untuk anak yang memperoleh skor 1007
5. Tes tes A dan B memiliki reliabilities dari ,90 dan ,80; korelasi antara tes A dan B adalah,50. Apa yang dimaksud dengan keandalan perbedaan antara nilai pada tes tes A dan B?

Jawaban

1. 76.118,%, 100.136
2. 8
3. 92, 60
4. 105, 95
5. .70

TAMBAHAN BACAAN

American Educational Research Association, American Psychological Association, & National Council on Pengukuran dalam Pendidikan. (1985). Standar ' pengujian andpsicholngicul pendidikan Washington, DC: American Psycho - Asosiasi logis. Coates, T., & Thoresen, C. (1978). Menggunakan teori generalisasi perilaku pengamatan. Terapi perilaku, Q 605-613. Cronbach, L., Gleser, G., Nanda, H., & Rajaratnam, N. (1972). Yang tergantung - kemampuan pengukuran perilaku Teori generalisasi uf nf skor dan profil New York: Wiley.