**APLIKASI RASCH MODEL DALAM MENGEVALUASI *INTELLIGENZ STRUKTUR TEST***

 ***Yulinda Erma Suryani1, Ummu Hany Almasitoh2***

1,2Universitas Widya Dharma Klaten; Jl. Ki Hajar Dewantara, Klaten Utara, (0272)326000

1yulinda@unwidha.id, 2ummuhany@unwidha.id

**Abstract**

The meaning and example of *intelligence* test instrument is important thing for the psychological research. The result of intelligence measurement described with IQ (intelligence quotient) which is the ratio of mental age and chronological age range of individual. Meanwhile, the ratio score of IQ does not fulfill the main element of its measurement including basic quantity, interval equivalence and absolute zero point. Based on the reason stated, the researchers developed the intelligence measurement model by using the Rasch model. The basic concept of this model should be objective on measuring the social and education values, and should be qualified to the following criterion: 1) provide the linear measurement with the same interval, 2) conducting the the appropriate estimation process, 3) finding the misfit and outlier items, 4) anticipating the missing data, 5) providing the replicable measurement and independently used as the research parameter. It is known that RASCH model has those qualifications mentioned as the intelligence measurement. Its quality is equal with the measurement that used on measuring physical dimension and field. By using RASCH model, the result of log odds unit is a scale which is on the same linearity and interval produced from the ratio data (odds ratio). The result of data analysis of IST test instrument describes the quality of IST test is categorized into **good** in general. From 176 test items there is only 1 item which identified as item 155 (WU19) categorized into **poor** and it should be eliminated as the consequence. Based on DIF analysis, it can be described that there are 28 items tend to take a side on a certain gender which means those 28 items need to be revised.

Key words: Intelligence, RASCH model

**Abstrak.** Makna dan keterandalan instrumen tes intelinsi sangat penting bagai penelitian psikologi. Hasil pengukuran inteligensi dinyatakan dengan IQ (*intelligence quotient*) yang merupakan merupakan rasio usia mental seseorang dan usia kronologis. Namun nilai rasio pada IQ belum memenuhi elemen penting dalam pengukuran yaitu besaran, kesetaraan interval dan nilai nol mutlak. Oleh karena itu melalui penelitian ini akan dikembangkan model pengukuran inteligensi yang akan menggunakan Rasch model. Konsep pengukuran yang objektif dalam ilmu-ilmu sosial dan penilaian pendidikan harus mempunyai lima kriteria, yaitu: 1) Memberikan ukuran yang linear dengan interval yang sama; 2) Melakukan proses estimasi yang tepat; 3) Menemukan aitem yang tidak tepat (*misfits*) atau tidak umum (*outliers*); 4) Mengatasi data yang hilang; 5) Menghasilkan pengukuran yang *replicable* (independen dari parameter yang diteliti). Kelima syarat pengukuran tersebut, sejauh ini hanya rasch model lah yang bisa memenuhinya. Kualitas pengukuran inteligensi yang dilakukan dengan rasch model akan mempunyai kualitas yang sama seperti halnya pengukuran yang dilakukan dalam dimensi fisik dalam bidang fisika. Skala logit (*log odds unit*) yang dihasilkan dalam model rasch adalah skala dengan interval yang sama dan bersifat linear yang berasal dari data ratio (*odds ratio*). Berdasarkan hasil analisis yang telah dilakukan terhadap instrumen tes IST dapat diketahui bahwa secara umum kualitas tes IST termasuk dalam kategori baik. Dar 176 aitem tes IST hanya ada 1 aitem yang tidak bagus, yaitu aitem 155 (WU19) sehingga aitem 155 sebaiknya dibuang. Berdasarkan analisis DIF dapat diketahui bahwa ada 28 aitem yang memihak kepada salah satu jender saja, sehingga keduapuluhdelapan aitem tersebut sebainya direvisi.

*Kata kunci*: (Inteligensi, Rasch Model)

Tes adalah alat atau teknik pengukuran yang digunakan untuk mengukur perilaku atau membantu memahami dan memprediksi perilaku. Sebuat tes psikologi (*psycholigical*) atau tes pendidikan adalah seperangkat aitem yang dirancang untuk mengukur karakteristik manusia yang berhubungan dengan perilaku, baik itu perilaku overt (*overt behavior*) dan perilaku covert (*covert behavior*). Oleh karena itu, tes psikologi dan pendidikan mengukur perilaku di masa lampau dan di masa sekarang Beberapa tes dibuat sebagai upaya untuk memprediksi perilaku di masa depan, misalnya kesuksesan di perguruan tinggi atau program jenjang yang lebih tinggi. Akan tetapi tes yang dibuat tidak mungkin dapat memprediksi secara tepat kesuksesan di pergurun tinggi karena tes hanya mengukur sebuah sampel perilaku dan kesalahan selalu berkaitaan dengan proses pengambilan sampel. Walaupun skor tes tidak mengukur secara sempurna suatu perilaku atau karakteristik, namun sangat membantu dalam hal proses prediksi. Dengan menggunakan pemodelan Rasch kualitas pengukuran inteligensi yang dilakukan akan mempunyai kualitas yang sama seperti halnya pengukuran yang dilakukan dalam dimensi fisik dalam bidang fisika. Pengukuran inteligensi dengan model Rasch menghasilkan analisis statistik kesesuaian (*fit statistics*) yang memberikan informasi pada peneliti apakah data yang didapatkan memang secara ideal menggambarkan bahwa orang yang mempunyai abilitas tinggi memberikan pola jawaban terhadap aitem sesuai dengan tingkat kesulitannya. Hasil pengukuran inteligensi yang menggunakan pemodelan Rasch bisa memenuhi ketiga elemen penting dalam pengukuran.

Hasil pengukuran inteligensi dinyatakan dengan IQ (*intelligence quotient*) yang merupakan hasil perbandingan antara usia mental dengan usia kronologis yang kemudian hasilnya dikalikan seratus untuk menghilangkan pecahan. Lebih khususnya IQ merupakan rasio usia mental seseorang (yang ditentukan dari performanya pada skala inteligensi) dan usia kronologis. IQ merupakan nilai rasio yang diperkenalkan pada revisi Stanford-Binet tahun 1916 oleh Terman. Jika IQ merupakan skala rasio berarti skor IQ harus memiliki ketiga elemen penting dalam pengukuran yaitu: magnitude (besaran), kesetaraan interval, dan 0 (nol) mutlak. Hasil pengukuran inteligensi memang memiliki magnitude (besaran), semakin tinggi skor yang diperoleh individu dalam tes inteligensi itu artinya semakin tinggi IQ individu tersebut. Hal itu berarti hasil pengukuran inteligensi memiliki tingkat kekuatan. Namun, apakah hasil pengukuran inteligensi memiliki kesetaraan inverval dan 0 (nol) mutlak? Tentu saja tidak, karena perbedaan antara IQ 45 dan 50 tidak memiliki makna yang sama dengan perbedaan IQ 105 dan 110. Walaupun masing-masing skor tersebut memiliki perbedaan 5 poin (50-45=5, dan 110-105=5), angka 5 poin yang pertama tidak memiliki makna yang sama dengan 5 poin kedua. Jika IQ memiliki kesetaraan interval itu artinya perbedaan antara IQ 45 dan 50 seharusnya bermakna yang sama dengan perbedaan IQ 105 dan 110. Karena tidak ada kesetaraan interval berarti skor IQ bukanlah interval melainkan ordinal. Demikian pula dengan nilai 0 (nol) mutlak, jika hasil pengukuran inteligensi berkisar pada skala 0-10, apakah itu berarti individu tersebut tidak memiliki inteligensi? Tentu saja hal itu sulit untuk dijelaskan. Berarti hasil pengukuran inteligensi yang dilakukan selama ini belum memenuhi ketiga elemen penting dalam pengukuran karena baru satu yang baru terpenuhi yaitu besaran (magnitude). Hal itu bisa terjadi karena penyekoran hasil pengukuran inteligensi menggunakan *clasical test theory* (CTT) yang memiliki banyak keterbatasan. Tingkat kesulitan dan pembobotan butir soal sangat tergantung pada karakteristik kelompok sampel yang diuji. Skor mentah dan murni akan berubah-ubah tergantung tingkat kesulitan soal yang diberikan.

Untuk mengatasi kelemahan dari hasil pengukuran pengukuran inteligensi yang menggunakan CTT tersebut, maka sangat penting untuk dilakukan pengembangan model pengukuran inteligensi. Hasil pengukuran inteligensi yang tidak terikat pada sampel butir soal tertentu atau orang yang dipilih (*item free dan person free*) menyebabkan pengukuran inteligensi lebih tepat di lakukan. Sehingga hasil pengukuran inteligensi bisa menggambarkan kondisi sebenarnya. Menurut Mok dan Wright, 2004 (dalam Sumintono, 2013) konsep pengukuran yang objektif dalam ilmu-ilmu sosial dan penilaian pendidikan harus mempunyai lima kriteria, yaitu: 1) Memberikan ukuran yang linear dengan interval yang sama; 2) Melakukan proses estimasi yang tepat; 3) Menemukan aitem yang tidak tepat (*misfits*) atau tidak umum (*outliers*); 4). Mengatasi data yang hilang; dan 5) Menghasilkan pengukuran yang *replicable* (independen dari parameter yang diteliti). Dengan adanya kelima syarat yang harus dipenuhi dalam pengukuran, sejauh ini hanya rasch model lah yang bisa memenuhinya. Sehingga dengan menggunakan pemodelan Rasch kualitas pengukuran inteligensi yang dilakukan akan mempunyai kualitas yang sama seperti halnya pengukuran yang dilakukan dalam dimensi fisik dalam bidang fisika. Pengukuran inteligensi dengan model Rasch menghasilkan analisis statistik kesesuaian (*fit statistics*) yang memberikan informasi pada peneliti apakah data yang didapatkan memang secara ideal menggambarkan bahwa orang yang mempunyai abilitas tinggi memberikan pola jawaban terhadap aitem sesuai dengan tingkat kesulitannyaHasil pengukuran inteligensi yang menggunakan pemodelan Rasch bisa memenuhi ketiga elemen penting dalam pengukuran. Selain itu, uji kuantitatif instrument yang dilakukan dengan pemodelan Rasch tidak hanya mengukur indeks realibilitas (alpha Cronbach) tetapi analisis aitem dilakukan ke tingkat masing-masing aitem. Selain terhadap aitem, rasch model juga secara bersamaan menguji person (responden), dimana akan terlihat pola jawaban responden yang konsisten, maupun mengidentifikasi jawaban yang asal saja melainkan juga untuk menguji dimensionalitas dan deteksi adanya bias dari aitem pada instrumen inteligensi.

**Metode**

Penelitian ini merupakan penelitian kuantitatif deskriptif yang bertujuan untuk melakukan validasi pengukuran inteligensi dengan menggunakan rasch model. Metode pengumpulan data yang digunakan dalam penelitian ini adalah dengan menggunakan metode dokumentasi. Data dalam penelitian ini diperoleh dari mengutip respon peserta dalam mengerjakan *Intelligenz Struktur Test* (IST).

Validasi instrument tes *Intelligenz Struktur Test* yang dilakukan dalam menelitian ini meliput analisis reliabilitas dan validitas dengan menggunakan bantuan perangkat lunak Winsteps (Linacre, 2011). Data yang sesuai dengan Rasch model memiliki nilai jumlah kuadrat tengah (*mean square*) sebesar 1,0 dan nilai terstandarkan (Z-*standardized values*) sebesar 0,0. Pada analisis aitem atau responden parameter yang digunakan untuk menunjukkan kesesuaian aitem dan respondengan dengan model harus memenuhi tiga kriteria, yaitu : *Point Measure Correlation* (x) : *0.32 < x < 0.8; Outfit Mean Square* (y): *0.5 < y < 1.5,* da*n Outfit Z standard* (z): *-2.0 < z <+2.0.* Aitem dikatakan tidak sesuai (*misfit*) apabila aitem tersebut terlalu mudah (nilai logit terlalu negatif) atau terlalu sulit (nilai logit positif yang besar); atau nilai logit yang dihasilkan tidak memenuhi tiga kriteria tersebut di atas. Aitem-aitem yang tidak memenuhi syarat mengindikasikan bahwa aitem tersebut tidak mengukur ciri-sifat yang diinginkan (Sumintono & Widhiarso, 2013). Selain itu, Rasch model juga bisa menganalisis *Diferential Aitem Functioning* (DIF). Analisis DIF yang dilakukan dalam penelitian ini adalah berdasarkan gender. Kriteria yang digunakan dalam menganalisis adalah dengan menggunakan nilai PROB, jika prob< 0,05 itu artinya aitem tersebut mengandung bias (DIF).

**H a s i l**

*Intelligenz Struktur Test* (IST) merupakan salah satu tes inteligensi standar yang mempunyai kemampuan untuk mengukur taraf inteligensi seseorang. Tes ini dikembangkan oleh Rudolf Amthauer di Frankfurt, Jerman pada tahun 1953. Tes ini dipandang sebagai gestalt (menyeluruh), yang terdiri dari bagian- bagian yang saling berhubungan secara makna (struktur). Dimana struktur intelegensi tertentu meggambarkan pola kerja tertentu, sehingga akan cocok untuk profesi atau pekerjaan tertentu. Berdasarkan hal tersebut IST umum digunakan untuk memahami diri dan pengembangan pribadi, merencanakan pendidikan dan karier serta membantu pengambilan keputusan dalam hidup individu.

IST terdiri dari sembilan subtes yang keseluruhannya berjumlah 176 aitem. Masing-masing subtes memiliki batas waktu yang berbeda-beda masih banyak diadministrasikan dengan menggunakan manual. Evaluasi kualitas instrumen *Intelligenz Struktur Test* dilakukan dengan tujuan untuk memberikan informasi secara keseluruhan tentang kualitas pola respons responden secara keseluruhan, kualitas instrumen Tes IST maupun interaksi antara responden dengan intrumen penelitian yang digunakan.

Berdasarkan hasil analisis data dapat diketahui bahwa nilai rerata dalam mengerjakan soal tes IST adalah -0,96, hal ini berarti bahwa nilai rata-rata yang diperoleh lebih kecil dari pada nilai logit 0,0 yang menunjukkan bahwa abilitas (kemampuan) responden yang lebih kecil daripada tingkat kesulitan soal.

Berdasarkan hasil analisis reliabilitas instrumen tes IST yang diperoleh sebesar 0,97 dan reliabilitas person diperoleh sebesar 0,88 Hal ini menjunjukkan bahwa reliabilitas instrumen IST termasuk dalam kategori bagus sekali. Pada pemodelan Rasch selain diperoleh analisis reliabilitas instrumen maupun person juga diperoleh nilai alpha Cronbach, yang merupakan reliabilitas interaksi antara person dan butir-butir soala secara keseluruhan. Berdasarkan hasil analisis dapat diketahui bahwa nilai Alpha Cronbach yang peroleh sebesar 0,89 yang termasuk dalam kategori bagus.

Nilai rata-rata INFIT MNSQ dan OUTFIT MNSQ person secara berurutan adalah 1.00 dan 1.01. Hal ini menunjukkan bahwa nilai yang diperoleh adalah nilai yang ideal. Untuk nilai INFIT ZSTD dan OUTFIT ZSTD diperoleh sebesar -0,1 dan -0,1, nilai ini mendekati nilai ideal 0,0, yang berarti bahwa kualitasnya semakin baik. Demikian pula dengan nilai rata-rata INFIT MNSQ dan OUTFIT MNSQ aitem secara berurutan adalah 1.00 dan 1.01. Hal ini menunjukkan bahwa nilai yang diperoleh adalah nilai yang ideal. Untuk nilai INFIT ZSTD dan OUTFIT ZSTD diperoleh sebesar 0,0 dan 0,1, nilai ini mendekati nilai ideal 0,0, yang berarti bahwa kualitasnya semakin baik.

Berdasarkan hasil analisis tingkat kesulitan butir soal dapat dilihat bahwa soal no 76 (GE 16) , 93 (RA17), 96 (RA 20) merupakan soal yang paling sulit dengan nilai logit 4,25. Sedangkan soal 61 (GE1 ) merupakan soal yang paling mudah dengan nilai logit -0,615. Menurut pemodelan Rasch semakin tinggi nilai logit yang diperoleh makka hal itu menujukkan bahwa semakin sulit soal tersebut. Hasil analisis tingkat kesulitan butir soal dengan menggunakan Program Winstep lebih memudahkan penggunanya karena hasil analisis tingkat kesulitan butir soal telah diurutkan dari yang tingkat kesulitan tertinggi (GE16) sampai ke tingkat kesulitan terendah (GE1).

Berdasarkan analisis tingkat kesesuaian butir soal dapat diketahui bahwa aitem hanya ada 1 soal yang tidak memenuhi ketiga keriteria tersebut yaitu aitem 155 (WU19) sehingga soal nomor 155 dikatakan tidak fit sehingga harus dibuang. Sedangkan aitem yang harus di revisi sebanyak 10 aitem yaitu aitem 152 (WU16), aitem 150 (WU14), aitem 58(AN18), aitem 138(WU2), aitem 147(WU11), aitem 137(WU1),aitem 42(AN2), aitem 35(WA15), aitem 118(FA2), dan aitem 1(SE1).

Berdasarkan Analisis DIF dapat diketahui bahwa ada 28 aitem yang terjangkit DIF, yaitu aitem SE20, WA3, WA8, WA9, WA10, AN5, RA9 ZR4, FA3, FA6, FA11, FA13, WU2, WU6, WU9, WU11, WU12, WU15, WU20, ME2, ME5, ME6, ME7, ME8, ME10, ME13, ME14, dan ME16. Hal itu berarti bahwa ke duapuluh delapan aitem tersebut hanya menguntungkan salah satu jenis kelamin saja. 28 aitem tersebut perlu diperbaiki supaya tidak merugikan kelompok jender tertentu.

**Diskusi**

Setiap pengukuran selalu menghasilkan informasi mengenai hasil pengukuran. Informasi pengukuran ini tergantung dari hubungan antara tes dengan individu yang diukur. Informasi pengukuran dipengaruhi oleh variasi hasil yang didapatkan. Fungsi informasi dalam teori respon butir yang cocok dengan model dapat dibagi menjadi dua yaitu : fungsi informasi butir untuk butir dan fungsi informasi tes untuk tes. Pada teori tes klasik, analisis butir digunakan untuk memilih butir yang memadai bagi kelompok peserta tes, sedangkan dalam teori respon butir untuk memilih butir yang memadai yaitu dengan menggunakan fungsi informasi butir. Dalam teori tes klasik koefisien reliabilitas digunakan untuk menilai mutu perangkat tes yang terdiri atas sejumlah butir, sedangkan dalam teori respon butir untuk mencapai tujuan yang sama didapatkan melalui fungsi informasi tes. Fungsi informasi digunakan untuk menggambarkan kekuatan butir atau tes, seleksi butir tes dan membandingkan perangkat tes (Hambleton, 1991). Fungsi informasi menyatakan kekuatan atau sumbangan tes dalam mengungkap *latent trait* yang diukur oleh tes tersebut. level abilitas sangat rendah, informasi yang didapatkan dari pengukuran juga cukup rendah. Demikian pula dengan level abilitas sangat tinggi, informasi yang didapatkan dari pengukuran juga cukup rendah. Pada level abilitas sedang, informasi yang didapatkan oleh pengukuran sangat tinggi. Hal itu menunjukkan bahwa butir tes IST tersebut menghasilkan informasi yang optimal ketika diberikan pada subjek dengan abilitas sedang. Artinya, soal tes IST merupakan tes dengan tingkat kesulitan sedang. Oleh karena itu tes IST kurang tepat apabila digunakan untuk proses seleksi karyawan karena pada proses seleksi karyawan dibutuhkan tes yang soal-soalnya memiliki tingkat kesulitan yang tinggi, sehingga tes dapat memberikan informasi yang lebih detail mengenai kemampuan individu yang memiliki kemampuan tinggi. Berdasarkan hasil fungsi informasi tes dapat diketahui bahwa tes IST lebih cocok digunakan untuk tes penjurusan di SMA daripada untuk tes seleksi. Selain itu, fungsi informasi butir juga menunjukkan reliabilitas pengukuran. Semakin tinggi puncak informasi yang dicapai, semakin tinggi nilai reliabilitas pengukuran yang dilakukan (Sumintono & Widhiarso, 2015). Hal ini didukung dengan hasil analisis koefisien reliabilitas tes IST pada tabel 1 yaitu 0,97 dengan nilai measure atau logit = 0, menunjukkan bahwa tes IST cocok digunakan untuk tes penjurusan siswa SMA, sehingga informasi pengukuran yang didapatkan sangat tinggi.

**Simpulan**

Berdasarkan hasil analisis reliabilitas instrumen tes IST yang diperoleh sebesar 0,97 dan reliabilitas person diperoleh sebesar 0,88 Hal ini menjunjukkan bahwa reliabilitas instrumen IST termasuk dalam kategori bagus sekali. Berdasarkan hasil analisis dapat diketahui bahwa nilai Alpha Cronbach yang peroleh sebesar 0,89 yang termasuk dalam kategori bagus.

Berdasarkan hasil analisis tingkat kesulitan butir soal dapat dilihat bahwa soal no 76 (GE 16) , 93 (RA17), 96 (RA 20) merupakan soal yang paling sulit dengan nilai logit 4,25. Sedangkan soal 61 (GE1 ) merupakan soal yang paling mudah dengan nilai logit -0,615. Nilai rata-rata INFIT MNSQ dan OUTFIT MNSQ person secara berurutan adalah 1.00 dan 1.01. Hal ini menunjukkan bahwa nilai yang diperoleh adalah nilai yang ideal. Untuk nilai INFIT ZSTD dan OUTFIT ZSTD diperoleh sebesar -0,1 dan -0,1, nilai ini mendekati nilai ideal 0,0, yang berarti bahwa kualitasnya semakin baik. Demikian pula dengan nilai rata-rata INFIT MNSQ dan OUTFIT MNSQ aitem secara berurutan adalah 1.00 dan 1.01. Hal ini menunjukkan bahwa nilai yang diperoleh adalah nilai yang ideal. Untuk nilai INFIT ZSTD dan OUTFIT ZSTD diperoleh sebesar 0,0 dan 0,1, nilai ini mendekati nilai ideal 0,0, yang berarti bahwa kualitasnya semakin baik.

Berdasarkan analisis tingkat kesesuaian butir soal dapat diketahui bahwa aitem hanya ada 1 soal yang tidak memenuhi ketiga keriteria tersebut yaitu aitem 155 (WU19) sehingga soal nomor 155 dikatakan tidak fit sehingga harus dibuang. Sedangkan aitem yang harus di revisi sebanyak 10 aitem yaitu aitem 152 (WU16), aitem 150 (WU14), aitem 58(AN18), aitem 138(WU2), aitem 147(WU11), aitem 137(WU1),aitem 42(AN2), aitem 35(WA15), aitem 118(FA2), dan aitem 1(SE1).

Berdasarkan analisis DIF dapat diketahui bahwa ada 28 aitem yang terjangkit DIF. Hal itu berarti bahwa ke duapuluh delapan aitem tersebut hanya menguntungkan salah satu jenis kelamin saja. 28 aitem tersebut perlu diperbaiki supaya tidak merugikan kelompok jender tertentu. keduapuluhdelapan aitem tersebut sebaiknya direvisi. yaitu aitem SE20, WA3, WA8, WA9, WA10, AN5, RA9 ZR4, FA3, FA6, FA11, FA13, WU2, WU6, WU9, WU11, WU12, WU15, WU20, ME2, ME5, ME6, ME7, ME8, ME10, ME13, ME14, dan ME16.

**Daftar Pustaka**

Ancok, D., 1992. *Teknik Penyusunan Skala Pengukuran*. Yogyakarta: Pusat Penelitian Kependudukan UGM.

Azwar, S., 2000. *Penyusunan Skala Psikologi*. Yogyakarta: Pustaka Pelajar.

Azwar, S., 2000. *Reliabilitas dan Validitas.* Yogyakarta: Pustaka Pelajar.

Domino, George. 2006. *Psychological Testing an Indruction*. New York. Cambridge Univ.

Gregory, R J. 2013. *Tes Psikologi: Sejarah, Prinsip, dan Aplikasi (Terjemahan Amitya Kumara dan Mikael Seno)* Edisi 6. Erlangga

Groth, Marnat. 2003. *Psychological Assessment*. New Jersey. University of Minnesota

Hadi, S., 1989. *Metodologi Research*. Jilid II. Edisi kesepuluh. Yogyakarta: Yayasan Penerbit Fakultas Psikologi Universitas Gadjah Mada.

Hadi, S., 1991. *Analisis Butir untuk Instrumen Angket, Tes dab Skala Nilai dengan Basica*. Yogyakarta: Andi Offset.

Kaplan, R.M., Saccuzzo, D,P. 2012. *Pengukuran Psikologi (terjemahan Eko Prasetyo Widodo).* Jakarta: Salemba Humanika.

Kerlinger, F.N., 1996. *Asas-asas Penelitian Behavioral* (Terj.Landung R.Simatupang). Yogyakarta: Gadjah Mada University Press.

Mardapi, D. 2008. *Teknik Penyususnan Instrumen tes dan Non Tes*. Yogyakarta: Mitra Cendikia Press.

Mugiharjo, 1984. *Perbedaan Inteligensi Quotient (IQ) Antara Anak Yang Latar Belakang Pekerjaan Orang Tua Berbeda Pada Siswa SMPN II Purworejo*, Yogyakarta: IKIP Yogyakarta.

Periantalo, J. 2015. *Validasi Alat Ukur Psikologi:Aplikasi Praktis*. Yogyakarta:Pustaka Pelajar.

Sumintono, B., Widhiarso, W. 2013. *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial. Cimahi: Trim Komunikata.*

Sumintono, B., Widhiarso, W. 2015. *Aplikasi Pemodelan Rasch Pada Assesment Pendidikan*. Cimahi: Trim Komunikata.

Supratiknya, A. 2014. *Pengukuran Psikologis*. Yogyakarta: Universitas Sanata Dharma.

Suryabrata, S., 1999. *Pengembangan Alat Ukur Psikologis*. Yogyakarta: Andi Offset.