

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Tes adalah salah satu cara untuk melakukan pengukuran dan dapat mengumpulkan informasi karakteristik pada suatu objek. Menurut (Mardapi, 2008) tes adalah salah satu cara untuk menafsirkan lebih darinya kemampuan seseorang secara tidak langsung, yaitu melalui respon seseorang terhadap stimulus atau pernyataan. Tes juga dapat diartikan sebagai jumlah pernyataan yang harus diberikan tanggapan dengan tujuan untuk mengukur tingkat kemampuan seseorang atau mengungkap aspek tertentu dari orang yang diberi tes. Tes merupakan suatu prosedur yang sistematis untuk mengamati atau mendeskripsikan satu atau lebih karakteristik seseorang dengan menggunakan standar numerik atau sistem kategori (Cronbach, 1984). Tes juga dapat digunakan untuk mengukur banyaknya pengetahuan yang diperoleh individu dari suatu bahan pelajaran yang terbatas pada tingkat tertentu (Bruce, 1978). Tes merupakan salah satu alat pengukuran yang paling sering digunakan pada bidang pendidikan dan psikologi. Pada pelaksanaannya, tes seharusnya berasaskan objektif, transparan, akuntabel, dan tidak diskriminatif. Selain itu, tujuan tes juga adalah untuk menguji apakah suatu model teoritis tentang cara penggunaan skor tes untuk tujuan tertentu yang selama ini sering dipakai memang cukup handal dan dapat dipercaya (Umar, 2011).

Untuk mendapatkan instrumen tes yang berkualitas tinggi, selain dilakukan analisis secara teoritis, perlu juga dilakukan analisis tes secara empirik. Secara garis lebih darinya, analisis tes secara empirik ini dapat dibedakan menjadi dua yaitu dengan pendekatan teori tes klasik dan teori tes modern atau teori respon tes (*Item Response Theory, IRT*). Menurut Allen dan Yen (1979) menyebutkan bahwa teori tes klasik atau disebut teori skor murni klasik didasarkan pada suatu model aditif, yaitu skor amatan dan skor kesalahan pengukuran (*error score*). Skor amatan adalah penjumlahan dari skor sebenarnya sedangkan kesalahan pengukuran adalah kesalahan tidak sistematis atau kesalahan secara acak.

Terdapat beberapa asumsi dalam teori tes klasik. Asumsi yang pertama, skor kesalahan pengukuran tidak berinteraksi dengan skor sebenarnya. Kedua, skor

kesalahan tidak korelasi dengan skor sebenarnya dan skor kesalahan pada tes yang lain untuk peserta tes yang sama. Ketiga, rata-rata dari skor kesalahan ini sama dengan nol. Dari asumsi-asumsi inilah sehingga dapat dijadikan sebagai dasar pengembangan formula dalam menentukan suatu kesahihan atau keandalan pada suatu tes (Retnawati, 2016). Tes dikatakan valid, sah, atau andal apabila tes tersebut telah memenuhi kriteria uji tes. Kriteria untuk mengukur suatu tes adalah tes tersebut telah memenuhi tingkat indeks kesukarannya dan terdapat daya beda. Penyebutan kriteria tersebut berlaku pada teori tes klasik, pada teori tes modern biasa disebut dengan istilah *discriminant* dan *difficulty*.

Pada penggunaan tes secara klasik memiliki kelemahan. Menurut Croker & Algina (1986) kelemahan teori klasik pertama adalah tingkat kesukaran dan daya beda soal sangat tergantung pada sampel yang digunakan dalam analisis. Rata-rata tingkat kemampuan, rentang, dan sebaran kemampuan peserta didik yang dijadikan sampel dalam analisis sangat mempengaruhi skor statistik yang diperoleh. Kedua, skor yang diperoleh peserta didik dari suatu tes sangat terbatas pada tes yang digunakan sehingga tidak dapat digeneralisasikan di luar tes yang digunakan. Ketiga, reliabilitas tes pada kesejajaran perangkat tes sangat sukar untuk dipenuhi. Keempat, tidak memberikan landasan untuk menentukan bagaimana respon peserta didik apabila diberikan tes tertentu. Kelima, indeks kesalahan baku pengukuran diasumsikan sama untuk setiap peserta tes. Keenam, pengujian bias butir tes dan penyetaraan tes tidak bersifat praktis dan sukar untuk dilakukan. Apalagi mengenai penyetaraan yang bersifat vertikal. Sehingga, untuk mengatasi kelemahan dalam teori tes klasik tersebut maka dapat digunakan pendekatan teori tes modern atau IRT.

Pada IRT ada asumsi yang harus dipenuhi yaitu asumsi unidimensi. Secara keseluruhan tes, asumsi unidimensi artinya mengukur satu kemampuan yang sama (Sheng dan Wikle, 2007). Folk dan Green (1989) berpendapat bahwa yang terjadi adalah banyaknya tes mengukur lebih dari satu kemampuan (multidimensi). *Multidimensional Item Response Theory* (MIRT) dapat diterapkan untuk mengukur kemampuan umum atau kemampuan psikologi tertentu pada peserta tes jika tes bersifat multidimensi (Segall, 2000). Dari pelanggaran asumsi unidimensi tersebut, maka tes yang berdimensi lebih dari satu dapat mengarah pada terjadinya bias pada

estimasi parameter dan menghasilkan estimasi parameter yang tidak tepat. Dengan demikian, diperlukan adanya pengembangan model MIRT.

Reckase (1997) menyatakan bahwa MIRT dapat digunakan untuk pemulihan tes dalam rangka memprediksi pembelajaran maupun estimasi kemampuan peserta didik. Pengujian kegunaan model MIRT yang diusulkan dilakukan untuk menganalisis kecocokan model (*goodness of fit*). Bock dan Aitkin, (1981); Reckase dan McKinley, (1982, 1983a, 1983b, 1984); Samejima, (1974); Thissen dan Steinberg (1984), Whitely (1980). Muraki dan Engelhard (1985) menyarankan untuk estimasi parameter butir tes dan kemampuan sebaiknya menggunakan model multidimensi karena model ini sangat tepat dan efektif dibandingkan dengan model unidimensional.

Analisis tes dan analisis kemampuan dengan MIRT ini biasanya menggunakan data yang bersifat dikotomi. Seperti penelitian yang dilakukan oleh Bock dan Aitkin (1981), Bock & Lieberman (1970), Lieberman (1970), McDonald (1985), Mulaik (1972), Rasch (1961), Reckase (1973), Sympson (1978), Whiteley (1980). Pada model MIRT juga terdapat dua model yaitu model kompensatori dan nonkompensatori. Spray, Davey, dan Reckase, (1990) menguraikan definisi dari model kompensatori adalah kemampuan tinggi peserta didik pada salah satu dimensi yang dikompensasikan pada kemampuan rendah di dimensi lain dalam kaitannya dengan probabilitas menjawab benar. Model non kompensatori adalah model yang tidak membolehkan kemampuan tinggi pada salah satu dimensi untuk memperoleh kompensasi pada kemampuan rendah di dimensi lainnya. Terkait dengan penelitian ini, maka model yang digunakan adalah model MIRT kompensatori dengan menggunakan data dikotomi.

Penggunaan model MIRT disini adalah bagian dari pengembangan dari model nonlinier. Menurut Yalcin (1995) dan McDonald (1962) model nonlinier adalah model tes yang membedakan kemampuan peserta didik, dimana tes yang mudah untuk peserta didik yang memiliki kemampuan tinggi dan tes yang sulit untuk peserta didik yang memiliki kemampuan rendah. Model nonlinier yang digunakan pada penelitian ini adalah model multidimensi tiga parameter logistik atau disingkat M3PL. Tiga parameter logistik yang dimaksud adalah parameter taraf sukar, daya beda, dan parameter kebetulan menjawab benar. Reckase, *et al*

(1991) juga menggunakan Model M3PL dengan penggunaan data dikotomi (biner). Data biner adalah data dengan dua kemungkinan jawaban benar atau salah (Gunther A and M. Hofler, 2006). Jadi, model logistik ini memiliki probabilitas jawaban benar dari setiap tes dan probabilitas jawaban salah.

Gagasan terkait pemodelan pada variabel laten (tidak dapat diukur secara langsung) dalam analisis data eksplorasi multivariate, dimana beberapa struktur data yang berdimensi rendah pada data yang berdimensi tinggi adalah suatu hal yang menarik, akan tetapi tidak dapat diamati secara langsung karena adanya kesalahan. Asumsi pada analisis faktor inilah sehingga McDonald membangun suatu hubungan analisis faktor model linier berdasarkan pada prinsip independensi lokal dan dinyatakan bahwa model linier merupakan kasus khusus IRT. Prinsip independensi di bangun oleh McDonald pada tahun 1967, 1980, 1981, 1982a, 1982b, 1985a, 1985b, 1986, 1999, 2001; McDonald & Mok, (1995).

Asumsi IRT model linier adalah respon peserta didik pada satu butir tes hanya mengukur satu kemampuan yang dominan. Asumsi lainnya adalah respon peserta dalam memberikan jawaban terhadap setiap tes bersifat independen. Artinya, bahwa respon satu peserta didik terhadap satu tes tidak di pengaruhi oleh peserta didik lain dalam memberikan respon terhadap tes. Selain itu, respon setiap peserta didik terhadap tes diasumsikan hanya bergantung pada satu parameter. Hal ini disebut monotonik.

Beberapa penelitian telah didedikasikan untuk memeriksa hubungan antara model IRT umum, misalnya model logistik dan fungsi ojaif normal, dan *nonlinier faktor analisis*, NLFA (Bartholomew, 1983; Goldstein & Wood, 1989; Knol & Berger, 1991; McDonald, 1967, 1989; Takane & De Leeuw, 1987). Muthen (1978, 1983, 1984) telah menunjukkan bahwa model umum digunakan IRT misalnya, model ojaif normal dua parameter, adalah kasus spesifik dari model analisis faktor umum untuk variabel kategori dengan beberapa indikator (yaitu, kategori respons). Menurut McDonald (1982b), model faktor umum Spearman menunjukkan bahwa model IRT adalah kasus khusus dari NLFA dan menyediakan kerangka umum yang mencakup model unidimensional/multidimensi, linier/nonlinier serta dikotomis dan polikotomis.

Model linier dalam penelitian ini sebagai penyeimbang dari model nonlinier. Model linier adalah bentuk hubungan antara variabel yang diamati dengan variabel laten yang tidak linier dalam parameter. Model linier mengasumsikan variabel yang diamati adalah fungsi linier dari variabel laten. Dengan asumsi model linier pada tes dapat membedakan kemampuan peserta didik. Misalnya peserta didik dengan kemampuan tinggi akan mudah dalam menjawab tes dengan benar dan sebaliknya peserta didik yang memiliki kemampuan rendah merasa sulit dalam menjawab tes. Namun, jika tesnya sangat sulit atau sangat mudah maka asumsi tersebut tidak terbukti/benar. Jadi, peneliti berkesimpulan bahwa estimasi parameter tes dan kemampuan dengan menggunakan model linier dan nonlinier sangatlah penting. Untuk membuktikan asumsi tes yang bersifat MIRT berdasarkan model linier dan nonlinier maka perlu dilakukan analisis faktor.

Terdapat hubungan yang sangat erat antara model Analisis Faktor (AF), IRT dan MIRT. Hubungan analisis faktor, IRT model linier, dan IRT model nonlinier adalah dilihat dari sejarah penggunaan analisis faktor pertama dikembangkan oleh Horst (1965). Horst mengembangkan model faktor yang menghasilkan data berbentuk matrix. Untuk menstandarkan variabel ketika faktor mengalami matrix data dikotomi dan memisahkan efek dari taraf sukar tes. Horst melakukan transformasi tes pada tes *error*. Dari kelemahan tersebut, maka Christoffersson (1975) dan Muthen (1978) mengembangkan analisis faktor sebagai model probabilitas yang menghubungkan antara tes dan kemampuan menggunakan model ojaif normal untuk mendapatkan parameter *threshold*. Parameter *threshold* adalah nilai yang menyimpang pada kurva normal atau skor dengan proporsi respon salah dalam menjawab tes. Di dalam MIRT menyebutkan parameter *threshold* adalah nilai taraf sukar tes. McDonald (1999) menyebutkan bahwa hubungan faktor tes dan dimensi kemampuan adalah independen dan dimensi lainnya yang tidak dapat diukur dibiarkan nonlinier. Pada tahun (1967) McDonald mengembangkan analisis faktor model linier dengan MIRT yakni menggunakan model polinomial. Model polinomial digunakan sebagai perkiraan model ojaif normal untuk mewakili respon peserta didik dengan tes. Bock dan Aitken (1981) memberikan gagasan tentang analisis faktor dan IRT untuk menghasilkan MIRT yakni dengan mendefinisikan model ojaif normal untuk kemampuan multidimensi yang mencakup karakteristik

tes dengan merepresentasikan kemampuan peserta dengan nilai taraf sukar dan daya beda pada tes. Meskipun memiliki cara yang sama untuk melihat hubungan faktor ke dimensi namun analisis faktor dan IRT memiliki fokus yang berbeda. Analisis faktor mencoba menentukan jumlah minimum faktor yang mereproduksi data dalam matriks respon tes sedangkan analisis IRT mencoba untuk memodelkan hubungan antara karakteristik kemampuan dengan tes. Lord (1980) menambahkan bahwa perlu mendeskripsikan tes dengan parameter tes dan kemampuan peserta didik sehingga dapat memprediksi secara probabilitas respon setiap peserta didik terhadap tes.

IRT sering didasarkan pada asumsi bahwa parameter yang menggambarkan peserta didik hanya pada satu dimensi (Lord dan Novick, 1968; Rasch, 1960). Namun asumsi unidimensional sering dilanggar. Sehingga perlu ditelaah kembali dan di tentukan konsekuensi dari pelanggaran asumsi unidimensional (Camili et al, 1995; Kirisci et al, 2001; Wang dan Wilson, 2005). Hal yang sama juga dilakukan oleh Fischer dan Molenaar (1995) mengembangkan model logistik linier. Kemudian Resckase (1972) dan Kelderman (1994) menggunakan fungsi penilaian untuk menerapkan model yang disebut testlets untuk tes dengan dua kategori skor. Testlets adalah seperangkat tes yang telah ditetapkan secara berurutan. Glas (1992) menggunakan fungsi penilaian untuk menunjukkan dimensi konten yang diukur oleh tes. Rijmen dan De Boeck (2005) menggunakan model fungsi penilaian untuk menunjukkan dimensi yang diukur oleh tes pada konteks pengujian adaptif. Lord & Novick (1968) mempresentasikan hubungan antara model IRT ojaif normal dengan model faktor. Kemudian Samejima (1974) mengembangkan model MIRT pertama yang diberi skor dikotomi maupun politomi yang memungkinkan untuk tes dengan skala kontinu. Samejima menggeneralisasikan model respon secara kontinu adalah probabilitas untuk memperoleh skor  $z$  lebih baik pada tes  $i$  dan  $bi$  adalah parameter kesulitan.

Analisis faktor dan MIRT memiliki formula statistik yang hampir sama ketika diterapkan pada matriks respon tes seperti yang dijelaskan oleh Bock dan Aitken (1981), Samejima (1974), dan McDonald (1967). Bahkan perangkat lunak untuk analisis faktor informasi lengkap (Bock et al, 1988) dapat digunakan untuk MIRT. Perbedaannya adalah analisis faktor dianggap sebagai teknik reduksi data dengan

tujuan untuk menemukan jumlah faktor terkecil yang mereproduksi matriks korelasi yang diamati, sedangkan MIRT adalah teknik untuk memodelkan respon kemampuan dan tes. Pada MIRT, matriks menunjukkan perbedaan rata-rata dan variansi skor tes karena memiliki hubungan langsung dengan karakteristik dari tes seperti taraf sukar tes dan daya beda sedangkan analisis faktor mempertimbangkan variansi dan kovariansi.

Berdasarkan hasil kaji dari beberapa jurnal terkait analisis faktor model nonlinier maka dideskripsikan beberapa yang terkait dengan penelitian ini diantaranya adalah Etezadi-Amoli dan McDonald (1983) tentang analisis faktor nonlinier yang dimana setiap tes adalah independen berdasarkan fungsi informasi tes, Mooijaart dan Bentler (1986) kaitannya dengan model nonlinier yang membandingkan model linier dengan beberapa alternatif model nonlinier lainnya dilihat dari fungsi informasi tes. Yalcin & Amemiya (2001) tentang analisis faktor nonlinier sebagai metode statistik yang dilihat dari kemungkinan adanya transformasi linier berdasarkan nilai RMSE dan relative bias di tinjau dari ukuran sampel kecil, dan Maydeu-Olivares (2005) yang membandingkan analisis faktor model linier dan IRT nonlinier berdasarkan fungsi informasi kemungkinan maximum. Dari beberapa kajian tersebut maka judul penelitian ini adalah analisis faktor model linier dan model nonlinier untuk estimasi parameter tes dan kemampuan dengan manipulasi data yang ditinjau berdasarkan nilai *Root Mean Square Error*, RMSE.

Model nonlinier merupakan model yang mengukur kemampuan yang berbeda dalam menjawab tes. Segall (2000) menguraikan bahwa model nonlinier dapat mengukur keakuratan parameter tes yang dipengaruhi oleh banyaknya peserta tes dan panjang tes (Cohen, Kane dan Kim, 2001; Heri Retnawati, 2008). Model nonlinier merupakan model yang mengukur kemampuan yang berbeda dalam menjawab tes.

Dari uraian mengenai model nonlinier tersebut bahwa model MIRT merupakan bagian dari model nonlinier maka peran penting MIRT dalam dunia pendidikan adalah berkaitan dengan dimensi dalam tes karena sangat mempengaruhi penskoran, analisis data dan laporan hasilnya (Abedi, 1997; Kahraman dan Thomson, 2011). Selanjutnya Ackerman (1994), Reckase dan

McKinley (1991) menguraikan bahwa tes terdiri atas beberapa subtes maka sangat perlu untuk diperhatikan kombinasi dari skornya sehingga dapat menginterpretasikan skor kompositnya.

Menurut Champlain dan Andre F (1995) bahwa estimasi parameter tes model IRT dapat menggunakan bebas ukuran sampel dan untuk estimasi kemampuan bebas dalam menggunakan panjang tes karena dapat memecahkan masalah dalam pengukuran. Matematika yang dikemas dalam soal cerita diperlukan kemampuan bahasa dan kemampuan matematika untuk mendeterminasikan jawaban benar (Reckase, 1997). Menurut Bolt dan Lall (2003), Ackerman, Gierl, dan Walker (2003), Thulber, Shinn, dan Smolkowski (2002) juga menunjukkan bahwa dalam penyelesaian permasalahan dalam matematika, paling tidak ada tiga dimensi yang termuat. Dalam penelitian model nonlinier ini menggunakan model multidimensi tiga parameter yakni; parameter daya beda (*slope*), parameter taraf sukar (*threshold*), dan parameter kebetulan betul (*pseudo guessing*). Setiap parameter diuji kecocokannya dengan 40 butir tes UN Matematika. Setiap tes dilihat korelasi berdasarkan panjang tes dan ukuran sampel yang telah dikelompokkan yakni; (1) perbandingan hasil estimasi antara model linier dan model nonlinier pada ukuran sampel 500 dengan panjang tes 20, 30, dan 40, (2) perbandingan hasil estimasi antara model linier dan model nonlinier pada ukuran sampel 1000 dengan panjang tes 20, 30, dan 40, (3) perbandingan hasil estimasi antara model linier dan model nonlinier pada ukuran sampel 1500 dengan panjang tes 20, 30, dan 40, (4) perbandingan hasil estimasi antara model linier dan model nonlinier pada panjang tes 20 dengan ukuran sampel 500, 1000, dan 1500, (5) perbandingan hasil estimasi antara model linier dan model nonlinier pada panjang tes 30 dengan ukuran sampel 500, 1000, dan 1500, (6) perbandingan hasil estimasi antara model linier dan model nonlinier pada panjang tes 40 dengan ukuran sampel 500, 1000, dan 1500.

## 1.2 Pembatasan Masalah Penelitian

Pembatasan masalah dalam penelitian ini antara lain:

- 1.1.1 Model linier dan model nonlinier diestimasi berdasarkan panjang tes, yakni: 20, 30, dan 40.

- 1.1.2 Model linier dan model nonlinier diestimasi berdasarkan ukuran sampel, yakni: 500, 1000, dan 1500.
- 1.1.3 Model linier dan model nonlinier diestimasi berdasarkan panjang tes 20 pada ukuran sampel 500, panjang tes 20 pada ukuran sampel 1000, dan panjang tes 20 pada ukuran sampel 1500.
- 1.1.4 Model linier dan model nonlinier diestimasi berdasarkan panjang tes 30 pada ukuran sampel 500, panjang tes 30 pada ukuran sampel 1000, dan panjang tes 30 pada ukuran sampel 1500.
- 1.1.5 Model linier dan model nonlinier diestimasi berdasarkan panjang tes 40 pada ukuran sampel 500, panjang tes 40 pada ukuran sampel 1000, dan panjang tes 40 pada ukuran sampel 1500.
- 1.1.6 Model nonlinier menggunakan model multidimensi tiga parameter logistik yakni parameter taraf sukar tes atau diskriminan (*threshold*), parameter daya beda (*slope/difficulty*) dan kebetulan betul (*pseudo guessing/asymptote*).
- 1.1.7 Membandingkan hasil analisis antara model linier dengan model nonlinier berdasarkan skor variansi dari hasil *Root Mean Square of Error* (RMSE).
- 1.1.8 Dimensi yang terukur pada mata pelajaran matematika (dimensi umum, dimensi spasial, dan dimensi numerik).

### 1.3 Rumusan Masalah

Berdasarkan latar belakang masalah dan batasan masalah, maka rumusan masalah penelitian adalah:

- 1.1.9 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 500 dengan panjang tes 20?
- 1.1.10 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 500 dengan panjang tes 30?
- 1.1.11 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 500 dengan panjang tes 40?

- 1.1.12 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 1000 dengan panjang tes 20?
- 1.1.13 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 1000 dengan panjang tes 30?
- 1.1.14 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 1000 dengan panjang tes 40?
- 1.1.15 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 1500 dengan panjang tes 20?
- 1.1.16 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 1500 dengan panjang tes 30?
- 1.1.17 Estimasi parameter pada model linier dan model nonlinier manakah yang lebih presisi dilihat dari variansi skor RMSE mendekati nol (0) pada ukuran sampel 1500 dengan panjang tes 40?

#### **1.4 Tujuan Penelitian**

Tujuan penelitian ini adalah:

- 1.1.18 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran sampel 500 dengan panjang tes 20.
- 1.1.19 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran sampel 500 dengan panjang tes 30.
- 1.1.20 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran 500 dengan panjang tes 40.

- 1.1.21 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran sampel 1000 dengan panjang tes 20.
- 1.1.22 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran sampel 1000 dengan panjang tes 30.
- 1.1.23 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran sampel 1000 dengan panjang tes 40.
- 1.1.24 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran sampel 1500 dengan panjang tes 20.
- 1.1.25 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran sampel 1500 dengan panjang tes 30.
- 1.1.26 Untuk mengetahui hasil estimasi parameter perbandingan antara model linier dan model nonlinier berdasarkan hasil presisi variansi dari skor RMSE ditinjau dari ukuran sampel 1500 dengan panjang tes 40.

### **1.5 State of The Art**

Penelitian ini memiliki kebaruan yakni pada: (a) estimasi parameter antara model linier dan model nonlinier pada tes Ujian Nasional; (b) hasil estimasi parameter yang mengukur tingkat presisi antara model linier dan model nonlinier berdasarkan variansi skor RMSE dengan ukuran sampel 500 dan panjang tes 20, 30, dan 40; ukuran sampel 1000 dan panjang tes 20, 30, dan 40; serta ukuran sampel 1500.

Kebaruan ini didasarkan pada hasil kajian beberapa jurnal terkait terdiri dari; McDonald (1967) tentang analisis faktor model linier dan model nonlinier dengan peskoran suatu dimensi dari sekelompok tes, dan setiap hubungan antara tes dan dimensi adalah independen dilihat dari fungsi informasi tes. Etezadi-Amoli dan McDonald (1983) tentang analisis faktor nonlinier yang dimana setiap tes adalah independen berdasarkan fungsi informasi tes, Mooijaart dan Bentler (1986)

kaitannya dengan model nonlinier yang membandingkan model linier dengan beberapa alternatif model nonlinier lainnya dilihat dari fungsi informasi tes. Yalcin & Amemiya (2001) tentang analisis faktor nonlinier sebagai metode statistik yang dilihat dari kemungkinan adanya transformasi linier berdasarkan nilai RMSE dan relative bias di tinjau dari ukuran sampel kecil, dan Maydeu-Olivares (2005) yang membandingkan analisis faktor model linier dan IRT nonlinier berdasarkan fungsi informasi kemungkinan maximum. Dari beberapa hasil kajian tersebut maka judul penelitian ini adalah analisis faktor model linier dan model nonlinier untuk estimasi parameter tes dan kemampuan dengan memanipulasi data yang ditinjau berdasarkan nilai *Root Mean Square Error*, RMSE.

### **1.6 Road Map Penelitian**

Penelitian tentang model linier dan model nonlinier untuk estimasi parameter tes dan kemampuan ini merupakan penelitian yang masih jarang dilakukan. Penggagas model nonlinier ini adalah Bartlett (1953) dan McDonald (1962).

Ada banyak artikel terkait tentang model linier dan model nonlinier untuk estimasi parameter tes dan kemampuan. Namun dalam kajian ini hanya beberapa yang dapat diuraikan diantaranya adalah McDonald (1962) membahas tentang gagasan dasar analisis faktor nonlinier dan memperkenalkan model nonlinier dalam faktor namun linier dalam pembebanan faktor, misalnya model dengan variabel yang diamati menjadi politomi dalam faktor kecuai untuk *error*. McDonald mengusulkan prosedur estimasi dua langkah. Pertama, model faktor orthogonal linier dipasang plot pencar dari skor faktor yang diperkirakan untuk mendeteksi kemungkinan nonlinier. Jika hubungan nonlinier signifikan maka plot fungsi mewakili hubungan estimasi.

Champlain dan Andre F (1999) tentang tinjauan analisis faktor nonlinier dan hubungannya dengan IRT secara statistik dengan memberikan gambaran singkat terkait hubungan antara IRT dan analisis faktor model nonlinier (nonlinier faktor analisis, NLFA) dan menguraikan tiga model NLFA, yang berfokus pada (1) pendekatan polinomial R. McDonald's dari model ojaif normal (1967, 1982); (2) model analitik faktor A. Christoffersson (1975) dan B. Muthen (1984) untuk variabel dikotomis; dan (3) analitik faktor informasi lengkap model (R. Bock dan

M. Aitken, 1981; R. Bock, R. Gibbons, dan E. Muraki, 1988). Item dari dua formulir Tes Masuk Sekolah Hukum (LSAT) dikalibrasi menggunakan ketiga model ini untuk menilai tingkat komparabilitas estimasi parameter IRT menggunakan prosedur ini. Temuan menunjukkan bahwa parameter perkiraan cenderung sangat mirip, terlepas dari prosedur kalibrasi. Penelitian tambahan harus dilakukan dengan jumlah formulir LSAT yang lebih besar sebelum mencapai kesimpulan apapun tentang komparabilitasnya prosedur kalibrasi.

Yalcin dan Amemiya (2001). Yang menulis tentang analisis faktor nonlinier sebagai metode statistika. Menyatakan menunjukkan bahwa kontribusi statistik ditinjau dari dua masalah yakni model linier dapat menjadi realistis bahkan sebagai pendekatan dalam banyak aplikasi dan sering tidak sesuai dengan data yang baik tanpa meningkatkan jumlah faktor. Sebagai model eksplorasi, model analisis faktor konvensional gagal untuk mengatasi struktur nonlinier yang mendasari data multivariate.

McDonald, 1967, 1980, 1981, 1982a, 1982b, 1985a, 1985b, 1986, 1999, 2001; McDonald & Mok, 1995). Dua pendekatan dapat digunakan untuk menghubungkan model nonlinier yang digunakan dalam teori respon item (IRT) dengan model linier yang digunakan dalam analisis faktor. Salah satu pendekatan adalah dengan menggunakan analisis harmonik (misalnya, McDonald, 1967, 1982a). Pendekatan kedua adalah dengan menggunakan fungsi link (misalnya, McDonald, 1999; Moustaki & Knott, 2000).

Meydeu-Olivares (2005) dalam artikelnya tentang IRT linier, IRT nonlinier, dan analisis faktor merupakan satu kesatuan yang terpadu. Fokus penelitiannya adalah hubungan penggunaan data biner (dikotomi) pada model IRT linier dan model faktor. Hasil menunjukkan bahwa dengan menggunakan data biner kedua model ini tidak selalu setara. Jadi, dalam menghubungkan model faktor dengan model respons item linier, perlu untuk mempertimbangkan masalah estimasi, khususnya penggunaan metode informasi terbatas versus informasi lengkap. Yang menarik dari temuan dari Meydeu-Olivares adalah model linier tidak menarik ketika fungsi respon ke tes tidak dibatasi antara nol dan satu. Jadi pada prinsipnya, untuk nilai rendah dari sifat laten kemungkinan mendukung tes mungkin negatif, sedangkan untuk nilai yang cukup tinggi dari sifat laten probabilitasnya mendukung

tes mungkin  $>1$ . Dari dua data yang digunakan pada model tes linier merupakan model yang tepat karena probabilitas sel berada dalam kisaran 0, 1.

Seiring berjalannya waktu maka penelitian model linier dan model nonlinier dikembangkan. Untuk di Indonesia pertama kali diteliti oleh Bastari (1999) tentang investigasi estimasi linier dan nonlinier multidimensi *graded response* data politomi, namun belum publish di jurnal internasional. Untuk itu, muncul ide untuk mencoba mengeksplorasi dan mendeskripsikan analisis faktor model linier dan model nonlinier menggunakan data dikotomi dalam penelitian ini. Penggunaan data yang berbentuk dikotomi ini karena menurut McDonald bahwa analisis faktor model nonlinier dengan mengukur model ojaif normal atau model logistik merupakan mitra dalam menggunakan data dikotomi. Dari pendapat tersebut maka penggunaan data dikotomi ini dilakukan. Tentunya, sangat berbeda antara politomi dan dikotomi baik dari segi analisis, pengembangan, dan penggunaannya. Penelitian ini menggunakan data dikotomi dan membandingkan dua *software* yang berbeda di setiap model. Penelitian ini difokuskan pada estimasi parameter model linier dan model nonlinier untuk estimasi parameter tes dan kemampuan. Analisis data dibangun dengan simulasi untuk memperoleh data yang sebenarnya (*true*) dan sebagai pembanding antara data *true* dan hasil perhitungan secara empirik.

Berdasarkan hasil kajian beberapa artikel, terdapat temuan dalam penelitian yang dilakukan oleh Dirk L. Knol dan Martijn P.F. Berger (2010) pertama, bahwa penggunaan satu dimensi menghasilkan nilai yang tidak signifikan maka disarankan untuk menggunakan dua dimensi. Penggunaan dua dimensi, menghasilkan pengujian dan tingkat keandalan yang lebih tinggi. Takane dan Leeuw, (1987) menggunakan dua dimensi dalam penelitiannya. Kedua, kumpulan data yang disimulasikan sangat sulit sehingga data yang dibangkitkan diperlukan untuk digeneralisasi. Namun Rose Marie Batley (1989) menegaskan bahwa penelitian dengan menggunakan dua parameter logistik menunjukkan bahwa parameter tes dan kemampuan memiliki korelasi yang homogen. Batley menyarankan bahwa perlu dilakukan penelitian lainnya dengan menggunakan tiga parameter logistik. Saran dari penelitian tersebut sehingga menjadi alasan pada penelitian ini untuk mengukur tiga parameter logistik pada tes multidimensi dan akan dilakukan *generate* data untuk menghasilkan data simulasi agar mendekati data riil.

Penelitian relevan yang telah dilakukan penulis antara lain; (1) *the estimates item parameter for multidimensional three parameter* (telah dipresentasikan dalam internasional konferensi yang diselenggarakan oleh IJERE Pascasarjana Universitas Negeri Jakarta (UNJ) pada Tahun 2018 namun baru dipublikasikan pada Tahun 2021, dan (2) *the exploration of nonlinier model was reviewed based on the length of the test* (telah dipublikasikan pada tanggal 18 oktober 2020 dan *accepted* pada tanggal 24 Desember 2020 di jurnal *Psychology and Education* terindex scopus Quartil 4 dan telah di verifikasi oleh GP3).

Rencana tahapan penelitian yang akan dilakukan adalah mengeksplorasi dan mendeskripsikan model linier dan model nonlinier menggunakan model multidimensional berdasarkan data dikotomi. Target luaran kedepan yang di telah disusun adalah publikasi disertasi dan membuat produk agar pembaca memiliki pegangan atau referensi tentang model nonlinier (model multidimensional).

