

EFEK URUTAN ITEM DI ERA MODERN: MODEL *ITEM POSITION EFFECT* DALAM *ITEM RESPONSE THEORY*

Dito Aryo Prabowo

Fakultas Psikologi, Universitas Diponegoro,
Prof. Mr. Sunario, Street, Tembalang, Semarang, Indonesia 50275

dito.aryo@live.undip.ac.id

Abstrak

Efek urutan item merupakan hipotesis pengaruh letak item pada performa tes. Tulisan ini bertujuan untuk membahas konsep dasar efek urutan item, sejarah penelitian, dan penerapannya dalam teori respon butir. Penelitian sebelumnya menunjukkan bahwa posisi item dalam tes memengaruhi tingkat kesulitan, diskriminasi, dan performa peserta tes. Model matematis yang dikembangkan untuk mempelajari efek ini meliputi model linear, nonlinear, dan kurvilinear. Dalam teori respon butir, efek posisi item dilihat melalui parameter *item difficulty* dan *differential item functioning*.

Kata kunci: Differential item functioning (DIF), item position effect, item response theory

Abstract

Item position effect hypothesizes the influence of item position in a test on performance. This article aims to explore the concept of item position effect, the history of research, and application in item response theory. Previous research shows that item position effect in a test influencing its item difficulty, item discrimination, and performance. Mathematical models developed to analyze this effect including linear, nonlinear, and curvilinear. In item response theory, item position effect utilized in item difficulty and differential item functioning.

Keywords: Differential item functioning (DIF), item position effect, item response theory

PENDAHULUAN

Item position effect merupakan isu di dalam asesmen, terutama asesmen berskala besar (*large scale assessment*) yang membutuhkan paket soal yang bervariasi (Albano, 2013). Isu yang dilandasi hipotesis bahwa pengerjaan individu dalam sebuah item dapat memengaruhi pengerjaan di item setelahnya, merupakan isu yang ramai dibahas di tahun 1950-1980 oleh peneliti psikometri (Leary & Dorans, 1985; Kang, 2014). Munculnya pengaruh urutan soal yang diberikan kepada peserta (selanjutnya disebut *test-taker*) didasari oleh konsep *context effect*, yang berasal dari *carry-over effect*, atau efek yang dihasilkan dari pengalaman sebelumnya pada perilaku yang diobservasi (Khorramdel & Frebort, 2011). *Context effect* sendiri dapat didefinisikan sesuai kesimpulan Albano (2012), yakni “*change in average performance across sets of items, or at the test level, given change in context or position*”.

Penelitian *item position effect* telah memberikan informasi penting terhadap penyusunan item di dalam sebuah tes. Leary dan Dorans (1985) misalnya, merekomendasikan item di *speed test* disusun dari yang termudah terlebih dahulu hingga yang paling sulit. Atau, bahwa posisi item lebih memengaruhi hasil tes *aptitude* dibandingkan tes *achievement*. Namun, di artikel yang sama, Leary dan Dorans (1985) menjelaskan bahwa penelitian *item position effect* dapat dilakukan dengan melibatkan *Item response theory* (teori respon butir, selanjutnya disebut

IRT), agar dapat melihat signifikansi yang lebih jelas dari *context effect* yang telah diidentifikasi.

Penelitian Kingston dan Dorans (1984) sebenarnya telah memberikan gambaran awal bagaimana kajian *item position effect* di dalam IRT. Mereka melihat pengaruh posisi item di dalam tes terhadap perilaku respon *test taker* pada level item. Hasil penelitian tersebut dianalisis menggunakan tiga estimasi parameter IRT, *item difficulty*, *item discrimination*, dan *pseudoguessing*. Selain itu perbandingan hasil *equating* antar tes yang diadministrasikan di dua kondisi juga menjadi salah satu analisis. Hasilnya, dari tiga subtes, terdapat satu subtes yang memiliki perbedaan *estimated item difficulty*.

Kajian IRT yang hadir pada konteks isu tes yang klasik, membuat penulis hendak mengulas tentang beberapa hasil kajian tersebut. Pada tulisan ini, secara singkat penulis akan memaparkan macam model matematika untuk estimasi *item position effect*.

METODE

Penelitian ini menggunakan teknik studi literatur untuk mengetahui isu-isu, area penelitian, dan teori berdasarkan sumber-sumber referensi yang relevan dengan topik yang diangkat (Ramdhani dkk., 2014). Pada penelitian ini, sumber-sumber referensi dipakai untuk mencari penjelasan tentang (a) definisi efek urutan item, (b) sejarah, dan (c) kaitan efek urutan item dan teori respons butir.

HASIL DAN PEMBAHASAN

I. *Item position effect*

Definisi *item position effect*

Item position effect sebagai sebuah istilah, memerlukan penjelasan yang lebih dalam pada kata yang digunakannya. Mengacu pada disertasi Albano (2012), *context effect* didefinisikan sebagai “*statistically significant change in item difficulty*”. Definisi ini mengikuti arti *context*, yaitu “... *reference to the items preceding a target item*” (Albano, 2012). Konteks yang dimaksud di dalam bahasan ini, mengacu pada item yang diberikan sebelum item yang sedang dibahas, dikarenakan ia memberikan konteks pada item tersebut. Kemudian, apabila ditemukan

Ringkasan sejarah penelitian *item position effect*

Penelitian *item position effect* dengan pemahaman teori tes klasik secara umum menggunakan eksperimen pemberian tes diberikan urutan tes yang berbeda. Kang (2014) menelusuri sejarah dari penelitian *item position effect*, dan menemukan dari literatur Leary dan Dorans (1985), ditemukan studi pertama tentang *context effect* di dalam tes dimulai oleh Mollenkopf di tahun 1950. Pada dekade penelitian yang mencari hubungan antara posisi item dan performa tes mulai bermunculan. Pada akhir 1960an hingga 1970an, efek interaksi antara penyusunan item (*item arrangement*, istilah yang sama untuk menyebut penyusunan item, seperti *item position* atau *item order*) dan karakteristik *test taker* mulai dilakukan. Penelitian di tahun 1980an mulai menggunakan pendekatan IRT.

Studi-studi awal *context effect* dimulai dari pengacakan di tingkat item dalam satu tes yang disajikan pada dua kelompok dan pengacakan di tingkat subtes (*section arrangement*), kemudian diuji perbedaan antar kelompok. Selain itu, Leary dan Dorans (1985) juga menemukan penyusunan item berdasarkan tingkat kesulitan. MacNicol (di dalam Leary & Dorans, 1985), melakukan penyusunan tersebut pada tes analogi verbal dengan partisipan sebanyak 1500 siswa SMA. Berdasarkan hasil eksperimen, ditemukan perbedaan performa tes

yang signifikan antara penyusunan item dari sulit ke mudah dan sebaliknya, dengan sulit ke mudah memiliki tingkat kesulitan yang lebih tinggi.

Selain penyusunan item, studi-studi berikutnya mengakomodasi efek interaksi yang berasal dari individu, seperti kecemasan, jenis kelamin, dan umur. Leary dan Dorman (1985) menjelaskan, meskipun variabel kecemasan *test taker* telah banyak diteliti, namun belum ditemukan kesimpulan bahwa kecemasan seseorang memengaruhi performa tes, dikarenakan hasil yang berbeda-beda.

Item position effect dan jenis tes

Leary dan Dorman (1985) meninjau studi-studi yang melihat efek penyusunan item terhadap performa tes berdasarkan jenis tes. Jenis tes yang telah diteliti ialah *power* dan *speed test*, serta *achievement* dan *aptitude*. Berdasarkan tinjauan yang dilakukan, terdapat beberapa temuan yang menarik. Misalnya, hasil studi pengaruh penyusunan item mudah-sulit di *speed test* akan menghasilkan skor yang lebih tinggi dibandingkan sulit-mudah. Selain itu, pada *power test*, penyusunan item dari mudah ke sulit tidak berpengaruh signifikan terhadap performa tes. Di sisi lain, dalam pembahasan *aptitude versus achievement test*, Leary dan Dorman (1985) menyimpulkan tes *aptitude* memberikan variasi terhadap performa *test taker* dibandingkan tes *achievement*.

Tes kepribadian juga menjadi salah satu studi dalam *item position effect*. Penelitian Schell dan Oswald (2013) dengan *Big Five personality inventory* dan Ortner (2004) dengan *Eysenck Personality Profiler*, menemukan tidak ada perbedaan rata-rata yang signifikan.

Pengaruh item position effect terhadap tes

Item position effect memberikan pengaruh terhadap performa *test taker* yang ditunjukkan dalam skor yang menjadi lebih rendah ketika diberikan soal yang lebih sulit terlebih dahulu, kemudian merasa lelah (*fatigue*), serta peningkatan *item difficulty* ketika soal sulit diberikan di awal (Leary & Dorman, 1985). Studi Le (2007) tentang posisi item di PISA, sebuah asesmen pendidikan berskala Internasional, dalam tes bagian Sains memiliki pengaruh terhadap kesulitan item dan kemampuan mendiskriminasikan *test taker* dengan kemampuan tinggi dan rendah.

II. Item response theory

Definisi item response theory

Item response theory atau IRT didasari oleh dua postulat dasar: (a) performa *test taker* pada item tes dapat diperkirakan atau dijelaskan oleh serangkaian faktor yang disebut *trait* atau kemampuan, (b) hubungan antara performa item *test taker* dan set *trait* di dalam performa item dapat digambarkan oleh fungsi bernama *item characteristic function* atau *item characteristic curve* (Hambleton, Swaminathan, & Rogers, 1991). Crocker dan Algina (2008) mendefinisikan IRT sebagai model matematika tentang bagaimana *test taker* dengan level kemampuan berbeda-beda untuk *trait* harus merespon sebuah item.

III. Penerapan item position effect di dalam item response theory

Macam-macam modelling

Terdapat beragam penelitian yang bertujuan untuk melihat model matematika yang tepat untuk menggambarkan *item position effect*. Kang (2014) dalam tinjauan literaturnya, membagi studi model *item position effect* menjadi tiga, yaitu linear, nonlinear, dan kurvilinear.

- (1) Studi model linear, menggunakan asumsi bahwa perubahan *item difficulty* terjadi secara linear, sehingga model yang dihasilkan dari asumsi ini lebih singkat dan membutuhkan parameter yang lebih sedikit. Oleh karena itu, di model ini *item position effect* diperlakukan sebagai *continuous variable*

- (2) Studi model nonlinear, menggunakan asumsi bahwa perubahan *item position effect* dapat dilihat sebagai *categorical variable*.
- (3) Studi model kurvilinear, melihat *position effect* dapat dimodel di dalam bentuk kuadrat (*quadratic terms*) yang diawali pemahaman bahwa kurvilinear adalah istilah polynomial untuk *continuous variable*. Model ini diawali dengan kemungkinan kasus dimana *item difficulty* meningkat pada awalnya, lalu menjadi stabil di tengah, dan berkurang di akhir. Kasus ini membuat estimasi efek dengan model linear tidak dapat memabahnya dengan baik.

Modeling using Item response theory framework

Debeer dan Jensen (2013) menggunakan pendekatan IRT untuk menjelaskan *item position effect*. Menurut mereka, *item position effect* di dalam IRT merupakan bagian dari DIF (*differential item functioning*). Hal ini disebabkan *item position effect* merupakan bias yang terjadi di dalam item tes yang berinteraksi dengan posisinya di dalam tes dan antar tes serupa. Studi *item position effect* dalam IRT, berfungsi dalam menjelaskan dampak posisi item terhadap karakteristik item. Pembahasan ini, sering diteliti dengan menggunakan pendekatan *two-step approach*, yang dimulai dengan menghitung estimasi *item difficulty* di setiap paket tes, dan dilanjutkan dengan melihat perbedaan *item difficulty* antar paket tes sebagai fungsi dari posisi item. Dalam model yang dijelaskan oleh Debeer dan Jensen (2013), asumsi efek linear, kuadrat, dan kubik akan memberikan estimasi yang paling tepat.

Modeling position effects on individual items. Debeer dan Jensen (2013) menjelaskan bahwa dalam melihat perbedaan *item difficulty* yang disebabkan oleh posisi item, dapat dilihat dengan DIF, sehingga *item position effect* diketahui dengan *one-step approach*, yang mana “*effect of item position can be distinguished from the general effects of person and item characteristics*”. Model ini dimulai dengan menggunakan logit, dengan notasi yang dipakai individu p ($p = 1, 2 \dots P$) ke item i ($I = 1, 2 \dots I$) dalam posisi k ($k = 1, 2 \dots K$) yang menjadi fungsi *latent trait* θ_p dan kesulitan β_{ik} .

$$\text{logit} [Y_{pik} = 1] = \theta_p - \beta_{ik} \quad (1)$$

Saat menggunakan parameter DIF, maka model dapat dikembangkan sesuai tujuan, yakni 1PL (*item difficulty*) dan 2PL (*item discrimination*). Dari studi literatur, ditemukan bahwa dalam tes *achievement*, semakin akhir letak sebuah item, makai akan menjadi lebih reliabel dan mampu mendiskriminasikan *test taker*.

Modeling item-position effects across items. Model selanjutnya menggunakan variabel letak item saja, dan tidak menggunakan interaksi antara isi dan posisi item seperti di model sebelumnya. Parameter posisi digunakan pada model ini, dengan notasi k (δ^{β}_{ik}).

$$\text{logit} [Y_{pik} = 1] = \theta_p - (\beta_{ik} + \delta^{\beta}_{ik}) \quad (2)$$

Modeling individual differences in position effects. Model dapat diperluas dengan memasukkan variabel perbedaan individu. Notasi yang ditambahkan adalah γ_p yang sebelumnya diartikan sebagai bobot linear dari posisi item, yang diubah menjadi bobot spesifik individu. Pada model ini, bobot diartikan sebagai *random effect* yang terdistribusi normal.

$$\text{logit} [Y_{pik} = 1] = \alpha_i [\theta_p - (\beta_{ik} + \gamma_p (k-1))] \quad (2)$$

Multilevel modelling

Albano (2013) menggunakan *framework* Hierarchical General Linear Model (HGLM) untuk membuat model *item position effect*. Model tersebut dikembangkan dengan menambahkan elemen parameter item dan individu, kemudian efek posisi diestimasi pada setiap item sambil mengontrol *item difficulty* dan *person ability*. Albano membandingkan tiga model yang divariasikan berdasarkan *item effect*.

$$\eta_{ij} = \gamma_{00} + \gamma_{q0} + u_{0j},$$

$$\eta_{ij} = \gamma_{00} + \gamma_{q0} + u_{0j} + \gamma_{N0}p_{ij},$$

$$\eta_{ij} = \gamma_{00} + \gamma_{q0} + u_{0j} + \gamma_{N0}p_{ij} + \gamma_{(N+q)0}p_{ij}.$$

Perubahan notasi yang berada dalam artikel ini berasal dari

$$\eta_{ij} = \beta_{0j} + \sum_{q=1}^{N-1} \beta_{qj} X_{qij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{qj} = \gamma_{q0},$$

Model ini menggabungkan model respon item di dalam individu dan *fixed item parameter* antar individu. Kemudian, model dimodifikasi dengan menambahkan notasi β_{Nj} sebagai *main effect* di level 1. Selain itu, efek interaksi antara posisi item dan item di model selanjutnya.

KESIMPULAN

Berdasarkan telaah literatur yang dilakukan, dapat diketahui bahwa meskipun *item position effect* diawali dari hipotesis yang berasal dari asumsi tes pada teori tes klasik (*classical test theory/CTT*), karena karakteristik tes yang *sample bound* (terikat oleh karakteristik peserta tes), membuat *item position effect* tidak banyak dibahas dalam penelitian terkini. Dalam studi di era modern yang dicirikan dengan Item Response Theory (IRT), *item position effect* masih dapat dikembangkan dengan mengembangkan model *item position effect*, melalui asumsi bahwa efek posisi merupakan bagian dari karakteristik tes yang hadir saat diletakkan di tempat tertentu.

Pengembangan model *item position effect* saat ini masih banyak dibahas pada *achievement test*, yang mana mempunyai *item difficulty* bervariasi, yang dalam penelitian-penelitian sebelumnya, lebih mudah untuk dibahas. Namun, pendekatan *item position effect* juga dapat dikembangkan ke dalam tes lain, seperti *power test*. Model-model yang disampaikan dalam telaah literatur menunjukkan keragaman asumsi peneliti-peneliti terdahulu atas *item position effect* dan asumsi yang mendasarinya.

REFERENSI

- Albano, A.D. (2012). Multilevel modeling of item position effects (disertasi doktor). The University of Minnesota, Minnesota, Amerika Serikat.
- Albano, A.D. (2013). Multilevel modeling of item position effects. *Journal of Educational Measurement*, 50(4). 408-426.
- Debeer, D., & Janssen, R. (2013). Modeling item-position effects within an IRT framework. *Journal of Educational Measurement*, 50(22), 164-185.
- Kang, C. (2014). *Linear and nonlinear modeling of item position effects* (tesis master). University of Nebraska, Nebraska, Amerika Serikat.

- Khorrarnadel, L., & Frebort, M. (2011). Context effects on test performance: What about test order? *European Journal of Psychological Assessment*, 27(2). 103–110. <http://doi.dx/10.1027/1015-5759/a000050>
- Kingston, N.M., & Dorans, N.J. (1982). The effect of the position of an item within a test on item responding behavior: An analysis based on item response theory (Laporan GREB-79-12bP (1982), RR-82-22). Diunduh dari [Educational Testing Service, <http://dx.doi.org/10.1002/j.2333-8504.1982.tb01308.x>].
- Le, L.T. (2007). Effects of item positions on their difficulty and discrimination: A study in PISA Science data across test language and countries. Makalah dipresentasikan pada 72nd Annual Meeting of the Psychometric Society, Tokyo.
- Leary, L.F., & Dorans, N.J. (1985). Implications for altering the context in which test items appear: A historical perspective on an immediate concern. *Review of Educational Research*, 55(3). 387-413.
- Ortner, T.M. (2004). On changing the position of items in personality questionnaires: Analysing effects of item sequence using IRT. *Psychology Science*, 46(4). 466-476.
- Schell, K.L., & Oswald, F.L. (2013). Item grouping and item randomization in personality measurement. *Personality and Individual Differences*, 55. 317-321. <http://dx.doi.org/10.1016/j.paid.2013.03.008>