

Game-based Assessment: The Development of Logical Reasoning Test

Wahyu Maulana Firdaus

Novita Sari

Aswin Januarsjaf

PT Global Talentlytica Indonesia, Jakarta

<https://doi.org/10.24071/suksma.v5i1.7650>

Naskah Masuk 14 Oktober 2023 Naskah Diterima 15 Desember 2023 Naskah Dipublikasikan 31 Januari 2024

Abstract. This study aims to test the validity of a game-based logical reasoning test tool called the Library Game. This tool consists of eight stimuli, each representing interrelated scenarios that follow a specific pattern to address problems. Data analysis results show that the eight stimuli in the Library Game are unidimensionally valid in measuring logical reasoning constructs ($n=284$, $\chi^2=30.742$, $df=21$, $p\text{-value}=0.078$, $RMSEA=0.040$, $CFI=0.996$, $TLI=0.995$). Furthermore, the researchers calibrated the items using Item Response Theory (IRT) 1 PL (parameter logistic). The results show that all items fit the established model. The level of item difficulty ranges from easy to moderate (-0.962 to 0.122). Therefore, this test can provide optimal information when administered to a population with an average level of logical reasoning. In conclusion, instead of conventional methods (question-answer format), valid assessment tools can be developed through innovative games.

Keywords: game-based assessment; logical reasoning; psychometric properties; item calibration

Pendahuluan

Di era kecerdasan buatan (*artificial intelligence*, AI) yang semakin maju saat ini, kebutuhan akan individu yang tidak hanya memiliki pengetahuan, tetapi juga keterampilan untuk bernalar secara logis menjadi semakin penting. Keterampilan ini tidak hanya relevan dalam pendidikan formal, tetapi juga dalam pengambilan keputusan sehari-hari, pemecahan masalah, dan peran profesional. Di samping itu, keterampilan bernalar secara logis juga menjadi kunci dalam menjawab tantangan-tantangan kompleks yang muncul di lingkungan profesional. Dalam beberapa literatur, kemampuan bernalar termasuk ke

Korespondensi Penulis

Wahyu Maulana, PT Global Talentlytica Indonesia, Jakarta

Email: wahyu@talentlytica.net

dalam salah satu aspek pada kemampuan mental umum (*general mental ability*, GMA) (Schneider & McGrew, 2018).

Di ranah industri, kemampuan bernalar logis memiliki dampak yang signifikan terhadap kinerja individu, termasuk pekerja (Salgado & Anderson, 2002; Schmidt & Hunter, 2004; Schmitt, 2014; Nye, Ma, & Wee, 2022). Kemampuan ini memberikan keunggulan penting dalam berbagai aspek profesional. Individu yang memiliki kemampuan penalaran logis yang kuat cenderung lebih efektif dalam pemecahan masalah, analisis data, dan pengambilan keputusan (Klauer & Phye, 2008; OECD, 2014). Mereka mampu mengidentifikasi pola, menyusun argumen yang kuat, dan menghindari kesalahan logika. Kemampuan ini juga membantu dalam merencanakan dan menjalankan tugas dengan lebih efisien, mengurangi risiko kesalahan, dan meningkatkan produktivitas (Green, Hill, Friday, & Friday, 2005). Selain itu, dalam lingkungan kerja yang semakin kompleks, individu dengan kemampuan bernalar logis yang baik cenderung lebih adaptif terhadap perubahan dan mampu menghadapi tantangan dengan lebih percaya diri (Niebaum & Bunge, 2014).

Mengakui urgensi dari keterampilan ini, riset ilmiah tentang pengembangan tes kemampuan penalaran logis telah menjadi fokus penting. Tes kemampuan penalaran logis dikembangkan untuk mengukur sejauh mana seseorang memiliki kemampuan untuk berpikir secara logis, menganalisis informasi dengan cermat, melihat keterkaitan antar suatu pola, dan mencapai kesimpulan yang tepat. Beberapa riset terdahulu telah berupaya mengembangkan tes yang bertujuan untuk mengukur tingkan penalaran logis individu. Misalnya *Test of Logical Thinking* (TOLT) mengukur aspek penalaran logis dengan kerangka lima model penalaran formal (Tobin & Capie, 1981). Hasil risetnya menunjukkan bahwa TOLT merupakan alat tes yang valid untuk mengukur penalaran pada siswa hingga perguruan tinggi. Contoh lainnya yaitu pengembangan tes kemampuan inteligensi umum berdasarkan kerangka teori Cattell-Horn-Carroll (CHC) yang didalamnya memuat tes penalaran (Schneider & McGrew, 2018). Menurutnya, dengan kerangka teori CHC dapat mengukur aspek kognitif secara lebih luas.

Dari contoh tes penalaran logis sebelumnya, keduanya dikembangkan dalam bentuk konvensional. Konvensional dalam hal ini diartikan bahwa tes berupa serangkaian pertanyaan yang perlu dikerjakan dan dijawab oleh peserta tes. Namun, seiring perkembangan zaman, tes-tes yang tidak konvensional mulai dikembangkan, seperti *Game-Based Assessment* (GBA) (Landers, Armstrong, Collmus, Mujcic, & Blaik, 2021). Landers, et al. (2021) mengembangkan setidaknya delapan jenis permainan yang mengukur kemampuan inteligensi umum. Tiga di antaranya yaitu, Numbubbles mengukur kemampuan kuantitatif, Resemble mengukur kemampuan bernalar, dan Proof It mengukur kemampuan membaca dan menulis. Tes seperti GBA dirancang menggunakan kerangka model permainan sehingga tes menjadi lebih atraktif. Sebuah tes yang atraktif sangat penting, khususnya di bidang industri untuk memotivasi para pekerja agar dapat mengerjakan tes hingga tuntas.

Meskipun tes berbasis permainan ini cukup efektif untuk menarik minat peserta untuk mengerjakannya (Molnar & Estrada, 2018), tetapi tantangan yang sering dihadapi yaitu terkait validitas permainan tersebut dalam mengukur konstruk yang dikehendaki. Lebih lanjut, aspek permainan bisa

mengalihkan fokus dari kemampuan yang diukur, sehingga skor yang dihasilkan mungkin lebih mencerminkan keterampilan bermain atau faktor lain, bukan kemampuan yang sebenarnya ingin diukur. Dengan demikian perlu dilakukan pengujian validitas konstruk untuk memastikan bahwa aspek-aspek permainan tidak mengganggu pengukuran yang valid dari konstruk yang diinginkan (Schneider & McGrew, 2018). Oleh karena itu, penelitian ini bertujuan untuk menguji validitas permainan yang dikembangkan oleh PT. Global Talentlytica Indonesia, salah satu perusahaan yang bergerak di bidang *online assessment platform*, berjudul “Library” dalam mengukur konstruk penalaran.

Metode Penelitian

Penelitian ini dilakukan dengan menyebarkan secara daring tautan pendaftaran keikutsertaan dalam riset pengujian permainan “Library”. Sebanyak 284 peserta (64% perempuan) berusia 18-52 tahun bersedia mengikuti riset ini. Pengambilan sampel dilakukan secara purposif dengan memilih sampel yang bersedia diambil datanya. Peserta yang bersedia mengikuti riset, diberikan tautan melalui surel untuk memainkan permainannya.

Permainan “Library” dirancang berdasarkan kerangka teori inteligensi umum, di mana salah satu faktornya yaitu *fluid reasoning* (*Gf*) (Schneider & McGrew, 2018). Dalam pengembangannya, tes ini diadaptasi dari tes pemecahan masalah yang disusun oleh *Organisation for Economic Co-operation and Development* (OECD) untuk studi *Programme for International Student Assessment* (PISA) (OECD, 2014). Pada asesmen tersebut, dijelaskan bahwa dalam pemecahan masalah, individu memerlukan kemampuan penalaran yang baik. Individu harus memahami setiap situasi yang ada, mengidentifikasi keterkaitan antar variabel, dan memperhitungkan tindakan yang dilakukan.

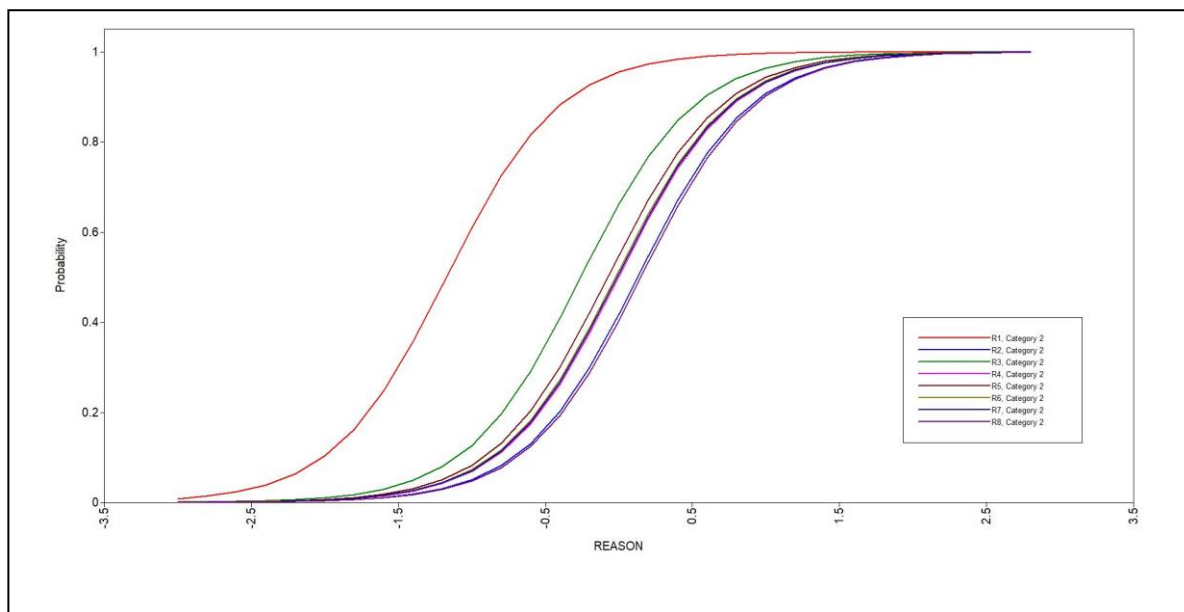
Dalam permainan “Library”, pemain, dalam hal ini peserta tes, diminta untuk memahami serangkaian situasi yang terdapat di perpustakaan. Pemain harus mampu melihat keterkaitan antar kondisi yang diberikan agar mendapatkan sebuah kesimpulan yang tepat. Pemain dikatakan berhasil atau memiliki kemampuan yang tinggi dalam permainan ini ketika ia mampu mengaitkan pola antar situasi dengan tepat, membuat langkah-langkah yang efisien, dan memanfaatkan waktu dengan baik.

Terdapat delapan stimulus dalam permainan yang akan digunakan sebagai item untuk dianalisis. Analisis yang digunakan yaitu analisis faktor konfirmatori (*Confirmatory Factor Analysis*; CFA) untuk menguji apakah delapan item tersebut dapat mengukur satu faktor penalaran logis (unidimensional). Selanjutnya setelah model satu faktor didapatkan, peneliti melanjutkannya dengan analisis response butir (Item Response Theory; IRT) 1 parameter logistik (1PL) untuk mendapatkan tingkat kesukaran item.

Hasil Penelitian dan Pembahasan

Berdasarkan hasil uji model fit CFA unidimensional, diketahui bahwa model delapan stimulus mengukur satu faktor sesuai (*fit*) dengan data empiris ($\chi^2=30.742$, $df=21$, $p\text{-value}=0.078$, $RMSEA=0.040$, $CFI=0.996$, $TLI=0.995$). Dari hasil tersebut dapat dikatakan bahwa permainan “Library” secara konstruk valid mengukur penalaran logis. Selanjutnya untuk mengkalibrasi parameter tingkat kesukaran tiap item, dilakukan analisis IRT 1PL. Dari hasil analisis didapatkan bahwa tingkat kesukaran soal merentang dari -0,962 sampai dengan 0,122, dengan tingkat daya beda yang paralel yaitu 0,821. Dengan demikian, permainan ini akan lebih optimal jika dimainkan oleh individu dengan kemampuan penalaran yang setara dengan rata-rata kebanyakan orang.

Gambar 1.

Kurva Karakteristik Item

Tabel 1.

Estimasi Parameter Item

Stimulus	Tingkat Kesukaran	Tingkat Daya Beda
Stimulus 1	-0.962	0.821
Stimulus 2	0.104	0.821
Stimulus 3	-0.213	0.821

Stimulus	Tingkat Kesukaran	Tingkat Daya Beda
Stimulus 4	-0.003	0.821
Stimulus 5	-0.063	0.821
Stimulus 6	-0.020	0.821
Stimulus 7	-0.011	0.821
Stimulus 8	0.122	0.821

Berdasarkan Tabel 1, item termudah yaitu pada Stimulus 1. Stimulus 1 menjadi item yang termudah karena stimulus tersebut merupakan acuan awal dalam permainan yang dapat membantu pemain dalam merespons stimulus-stimulus lainnya. Sedangkan tingkat kesukaran stimulus-stimulus lainnya berada pada rentang yang sama. Hal ini dikarenakan dalam permainan, terdapat satu stimulus sebagai acuan dan tujuh stimulus lainnya yang perlu dianalisis pola hubungannya.

Sebagai temuan awal dalam konteks asesmen berbasis permainan, hasil penelitian pada permainan "Library" ini sejalan dengan apa yang dikatakan Nunes (2012). Menurutnya, penalaran logis dapat diartikan sebagai bentuk berpikir di mana seseorang menghubungkan pola yang terdapat di antara beberapa premis/situasi untuk dapat mengambil kesimpulan dengan tepat. Dalam penalaran logis, individu menggunakan aturan-aturan dan prinsip-prinsip logika untuk membangun argumen yang konsisten dan valid. Kemampuan untuk berpikir secara logis memungkinkan seseorang untuk mencapai kesimpulan yang rasional, mengambil keputusan yang informasinya terbatas, dan berargumentasi dengan logika yang kuat (Nunes, 2012).

Kesimpulan dan Saran

Berdasarkan hasil uji model fit analisis faktor konfirmatori (CFA) unidimensional dan analisis *Item Response Theory* (IRT) 1PL, dapat disimpulkan bahwa permainan "Library" merupakan alat yang valid untuk mengukur kemampuan penalaran logis. Model dengan delapan stimulus dalam permainan ini secara baik mengukur faktor tunggal yang sesuai dengan data empiris. Hasil analisis IRT menunjukkan bahwa tingkat kesukaran tiap item dalam permainan bervariasi, dengan Stimulus 1 menjadi item paling mudah karena berfungsi sebagai acuan awal. Sementara tingkat kesukaran stimulus-stimulus lainnya berada dalam rentang yang sama, di mana pemain perlu mengaitkan pola antar stimulus tersebut untuk dapat menyelesaikan persoalan. Temuan ini sejalan dengan definisi penalaran logis, di mana individu harus menghubungkan pola antara beberapa premis atau situasi untuk

mengambil kesimpulan yang tepat (Nunes, 2012). Kemampuan berpikir logis ini sangat berharga dalam pengambilan keputusan yang berlandaskan pada logika yang kuat. Dengan demikian, permainan "Library" dapat digunakan secara efektif dalam konteks asesmen berbasis permainan untuk mengukur kemampuan penalaran logis individu.

Penelitian ini juga memiliki beberapa keterbatasan. Pertama, terkait dengan variasi item yang terbatas membuat permainan akan lebih mudah dikenali oleh pemain sehingga hal tersebut dapat mempengaruhi validitas skor yang dihasilkan. Untuk itu, perlu dilakukan pengembangan bank soal agar pola stimulus dalam permainan semakin bervariasi. Kedua, pola hubungan antar stimulus yang masih tergolong sederhana menjadikan item berada pada rentang tingkat kesukaran yang sedang dan rendah. Hal ini akan membatasi permainan ketika akan digunakan oleh orang dengan kemampuan kognitif yang tinggi. Oleh karenanya, selain pengembangan bank soal untuk variasi pola stimulus, perlu juga dipertimbangkan tingkat kesukarannya.

Daftar Acuan

- Green, A. L., Hill, A. Y., Friday, E., & Friday, S. S. (2005). The use of multiple intelligences to enhance team productivity. *Management Decision*, 43(3), 349–359.
doi:10.1108/00251740510589742
- Klauer, K., & Phye, G. (2008). Inductive reasoning: a training approach. *Review of Educational Research*, 78(1), 85–123.
- Landers, R., Armstrong, M., Collmus, A., Mujcic, S., & Blaik, J. (2021). Theory-driven game-based assessment of general cognitive ability: design theory, measurement, prediction of performance, and test fairness. *Journal of Applied Psychology*, 107. doi:10.1037/apl0000954
- Molnar, A., & Estrada, J. G. (2018). A comparative study of in-game and out-game assessment for storyline-based games. *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, pp. 138-142.
- Niebaum, J. C., & Bunge, S. A. (2014). Your brain is like a muscle: use it and make it strong. *Frontiers for Young Minds*, 1-6. doi:10.3389/frym.2014.00005
- Nunes, T. (2012). Logical reasoning and learning. *Encyclopedia of the Sciences of Learning*, 2066–2069. doi:10.1007/978-1-4419-1428-6_790
- Nye, C. D., Ma, J., & Wee, S. (2022). Cognitive ability and job performance: meta-analytic evidence for the validity of narrow cognitive abilities. *Journal of Business and Psychology*, 37, 1119-1139. doi:10.1007/s10869-022-09796-1
- OECD. (2014). *Pisa 2012 results: creative problem solving: students' skills in tackling real-life problems*. OECD Publishing. doi:10.1787/9789264208070-en
- Salgado, J., & Anderson, N. (2002). Cognitive and gma testing in the european community: issues

- and evidence. *Human Performance*, 15(1-2), 75-96. doi:10.1080/08959285.2002.9668084
- Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86(1), 162-173. doi:10.1037/0022-3514.86.1.162
- Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *The Annual Review of Organizational Psychology and Organizational Behavior*, 1, 45-65. doi:10.1146/annurev-orgpsych-031413-091255
- Schneider, W. J., & McGrew, K. S. (2018). The cattell–horn–carroll theory of cognitive abilities. In d. P. Flanagan, & e. M. Mcdonough. *Contemporary intellectual assessment: Theories, tests, and issues*. The Guilford Press.
- Tobin, K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. *Educational and Psychological Measurement*, 41(2), 413-423. doi:10.1177/001316448104100220