



Edukatif: Jurnal Ilmu Pendidikan

Volume 6 Nomor 4 Agustus 2024 Halaman 3975 - 3989

<https://edukatif.org/index.php/edukatif/index>

Differential Item Functioning (DIF): Sebuah Analisis Bibliometrik

Suwanda Priyadi

Fakultas Psikologi, Universitas Muhammadiyah Surakarta, Indonesia

e-mail : sp822@ums.ac.id

Abstrak

Differential item functioning (DIF) telah menjadi topik penelitian populer dalam beberapa tahun terakhir. Keberadaan DIF memiliki implikasi sosial yang besar dalam pengetesan, tidak hanya bagi peserta tes secara individual, namun juga pada tingkat penyusunan dan evaluasi kebijakan publik. Penelitian ini bertujuan untuk memberikan tinjauan kritis terhadap perkembangan publikasi ilmiah DIF menggunakan analisis bibliometrik. Data yang digunakan adalah 4.056 publikasi ilmiah DIF dalam basis data Scopus. Hasil analisis menunjukkan bahwa Amerika Serikat memiliki peran sentral dalam publikasi ilmiah DIF. Hal ini terlihat dari jumlah publikasi, jumlah afiliasi, dan jumlah kutipan yang didapatkan. Amerika Serikat juga menduduki posisi tertinggi dalam peta kolaborasi antarnegara. Sementara itu, analisis kemunculan kata kunci dan peta tematik menemukan dua tema utama dalam publikasi ilmiah DIF. Pertama, tema *niche* yang mencakup tema-tema seperti *Rasch analysis*, *psychometric properties*, *validity*, *quality of life*, *measurement*, *reliability*, dan *patient-reported outcomes*. Kedua, tema dasar yang mencakup tema-tema seperti *differential item functioning*, *item response theory*, *measurement invariance*, *factor analysis*, *assessment*, *depression*, dan *gender*. Hasil ini dapat menjadi referensi publikasi ilmiah yang berkaitan DIF di kemudian hari.

Kata Kunci: DIF, differential item functioning, analisis bibliometrik.

Abstract

Differential item functioning (DIF) has become a popular research topic in recent years. The presence of DIF has significant social implications for testing, not only for individual test takers but also for the development and evaluation of public policy. The purpose of this study is to conduct a critical review of DIF scientific publication development using bibliometric analysis. Data for the study came from 4,056 DIF scientific publications in the Scopus database. The analysis revealed that the United States plays a significant role in DIF scientific publications. This is demonstrated by the number of publications, affiliations, and citations obtained. The United States also leads in terms of international collaboration. Meanwhile, analyzing keyword occurrences and thematic maps revealed two major themes in DIF scientific publications. First, consider the niche theme, which covers topics such as *Rasch analysis*, *psychometric properties*, *validity*, *quality of life*, *measurement*, *reliability*, and *patient-reported outcomes*. The second theme is the basic theme, which covers topics such as *differential item functioning*, *item response theory*, *measurement invariance*, *factor analysis*, *assessment*, *depression*, and *gender*. The findings can be used as a reference in future scientific publications about DIF.

Keywords: DIF, differential item functioning, bibliometric analysis.

Copyright (c) 2024 Suwanda Priyadi

✉ Corresponding author :

Email : sp822@ums.ac.id

DOI : <https://doi.org/10.31004/edukatif.v6i4.7463>

ISSN 2656-8063 (Media Cetak)

ISSN 2656-8071 (Media Online)

PENDAHULUAN

Beberapa tahun belakangan, terjadi peningkatan minat penelitian *differential item functioning* atau biasa disebut DIF (Özmercan, 2023). Peningkatan minat terhadap DIF disebabkan oleh besarnya implikasi sosial yang mungkin ditimbulkan oleh keberadaan DIF dalam suatu pengetesan (Berrío dkk., 2020). Keberadaan DIF dapat membuat perbandingan skor antarkelompok menjadi tidak bermakna, membuat hasil tes menjadi tidak valid, dan menunjukkan bahwa instrumen tersebut tidak memberikan ukuran yang memadai terhadap konstruk yang ingin diungkap. Hal ini memiliki implikasi sosial yang tidak hanya bagi peserta tes secara individual, namun juga pada tingkat penyusunan dan evaluasi kebijakan secara luas.

Penelitian DIF pertama kali dilakukan pada tahun 1960an sebagai tanggapan atas kekhawatiran masyarakat bahwa tes kemampuan kognitif tidak adil bagi peserta ujian minoritas (Angoff, 1993). Kekhawatiran tersebut muncul karena tes kemampuan kognitif berisi butir soal yang berada di luar ranah budaya mereka, sehingga siswa minoritas memiliki sedikit kesempatan untuk belajar. Tujuan yang lebih spesifik dari penelitian ini adalah mengidentifikasi butir soal yang bias terhadap siswa minoritas dan menghapusnya dari tes. Awalnya, penelitian yang dilakukan untuk mengidentifikasi butir soal yang “bias” menggunakan istilah bias butir soal (*item bias*). Namun, penggunaan kata *bias* memiliki konflik semantik karena dapat dimengerti secara statistik dan sosial.

DIF adalah indikator statistik yang menunjukkan sejauh mana dua peserta tes yang memiliki tingkat kemampuan yang sama, namun berasal dari kelompok yang berbeda, memiliki probabilitas yang tidak sama dalam menjawab benar suatu butir soal (American Educational Research Association dkk., 2014). Gómez-Benito dkk. (2018) berpendapat bahwa DIF adalah bagian dari validitas karena penelitian DIF akan mengevaluasi faktor-faktor yang relevan dan tidak relevan dengan konstruk yang diukur serta konsekuensi negatif yang mungkin muncul dari keberadaan DIF. DIF tidak hanya terjadi pada tes yang mengungkap domain kognitif, namun juga terjadi pada skala psikologis yang mengungkap domain nonkognitif (Mumpuni dkk., 2023).

Perbedaan skor antarpeserta dalam suatu pengetesan perlu dievaluasi terlebih dahulu sebelum menjadi dasar pengambilan keputusan (Berrío dkk., 2020). Jika hal tersebut hanya mencerminkan perbedaan tingkat keterampilan, dengan kata lain, bahwa peserta tes atau kelompok peserta tes tertentu benar-benar lebih atau kurang mampu dalam aspek yang diukur oleh suatu soal, maka hal tersebut adalah dampak (*impact*). Namun, jika dua peserta ujian dengan tingkat kemampuan yang sama berbeda dalam hal kemungkinan mereka menjawab benar suatu soal, maka perbedaan tersebut adalah DIF. Jika terjadi DIF, maka peneliti perlu menganalisis lebih lanjut dengan memeriksa alasan mengapa butir soal tersebut menyebabkan tanggapan berbeda pada peserta atau kelompok yang berbeda. Hal ini dilakukan dengan mempertimbangkan tujuan tes, pengalaman peserta tes, informasi statistik hasil analisis butir soal atau kategori respons, dan aspek lainnya.

Beberapa penelitian telah melakukan tinjauan kritis terhadap perkembangan penelitian DIF (Berrío dkk., 2020; Özmercan, 2023; Rouquette dkk., 2019). Tinjauan tersebut penting untuk memberikan gambaran singkat mengenai perkembangan penelitian DIF serta rekomendasi bagi para peneliti yang tertarik dengan DIF. Misalnya, Rouquette dkk. (2019) meninjau praktik pendeteksian DIF skala komposit dalam penelitian bidang kesehatan. Berrío dkk. (2020) meninjau penggunaan data simulasi dalam pengujian metode pendeteksian DIF. Sementara itu, Özmercan (2023) meninjau literatur publikasi ilmiah DIF secara lebih umum. Tinjauan literatur yang dilakukan melibatkan basis data Medline dan *web of science* (WoS).

Perbandingan basis data Scopus dan WoS menunjukkan bahwa sebagian besar dokumen yang tercakup dalam Scopus tidak tercakup dalam WoS (Visser dkk., 2021). Beberapa dokumen tersebut adalah dokumen dengan jumlah referensi dan kutipan yang cukup banyak. Dokumen tersebut adalah publikasi artikel jurnal dan prosiding konferensi. Padahal dokumen tersebut penting untuk mengidentifikasi perkembangan publikasi dan topik penelitian dalam suatu disiplin ilmu. Sepengetahuan peneliti, belum ada penelitian yang

menganalisis publikasi ilmiah DIF menggunakan basis data Scopus. Oleh karena itu, penelitian ini meninjau perkembangan publikasi ilmiah DIF menggunakan basis data Scopus.

Tujuan penelitian ini adalah memberikan gambaran perkembangan publikasi ilmiah DIF, mengakui kontribusi ilmuwan di bidangnya, dan mengeksplorasi jaringan dan kerja sama publikasi antarlembaga dan negara afiliasi peneliti. Hal ini akan melengkapi hasil penelitian sebelumnya yang telah menunjukkan perkembangan publikasi DIF dalam bidang pendidikan, psikologi, psikiatri, kualitas hidup, dan kedokteran (Rouquette dkk., 2019). Secara khusus, penelitian ini akan menjawab pertanyaan berikut:

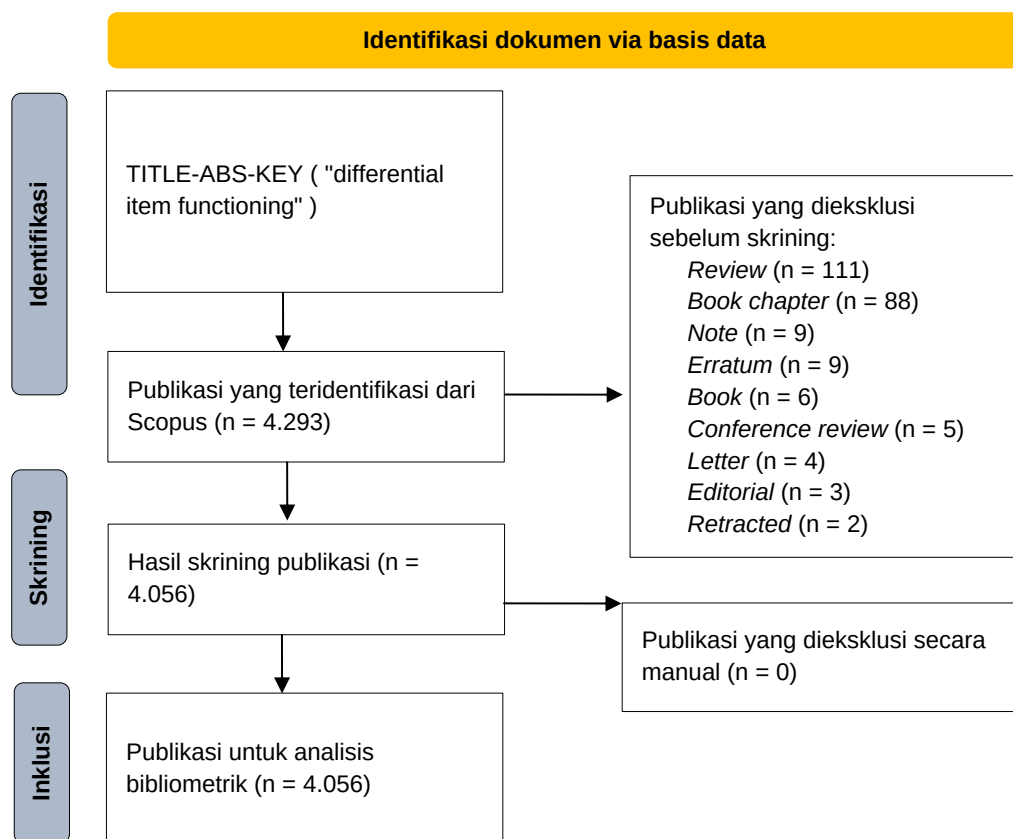
Q1. Bagaimana perkembangan publikasi ilmiah DIF?

Q2. Apa saja tema yang muncul dalam publikasi ilmiah DIF?

METODE

Penelitian ini menggunakan metode analisis bibliometrik. Analisis bibliometrik memungkinkan peneliti untuk memperoleh sudut pandang yang komprehensif mengenai topik penelitian yang dipilih dalam jangka waktu tertentu. Pelaksanaan analisis bibliometrik berpedoman pada Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page dkk., 2021).

Penelitian ini menggunakan basis data Scopus sebagai sumber data analisis bibliometrik. Proses identifikasi publikasi ilmiah dilakukan pada tanggal 11 Juli 2024. Pencarian pertama menghasilkan 4.293 publikasi. Kemudian, peneliti menetapkan artikel yang melalui telaah mitra bestari dan prosiding konferensi sebagai parameter pencarian untuk memastikan kualitas publikasi. Melalui PRISMA peneliti menetapkan kata kunci pencarian, identifikasi, dan skrining publikasi ilmiah. Model PRISMA dikembangkan secara khusus untuk membantu penulis dalam meningkatkan kualitas pelaporan pencarian perpustakaan dan secara efektif merangkum bukti-bukti yang ada melalui pendekatan langkah demi langkah yang sistematis dan transparan (Gambar 1).



Gambar 1. Diagram Alir PRISMA

Penelitian ini menggunakan paket-R bibliometrix (Aria & Cuccurullo, 2017), sebuah perangkat lunak sumber terbuka yang memfasilitasi penelitian bibliometrik dan pemetaan publikasi ilmiah. Teknologi pemetaan yang ditawarkan oleh bibliometrix terbukti menjadi alat yang berharga untuk melakukan analisis bibliometrik. Analisis bibliometrik telah mendapat pengakuan sebagai pendekatan berharga yang mempertimbangkan kontribusi intelektual, struktur sosial, dan kerangka konseptual pengetahuan di berbagai disiplin ilmu (Sari & Rahmawati, 2023; Supriani dkk., 2022; Supriyadi dkk., 2022; Zupic & Čater, 2015). Versi terbaru bibliometrix dilengkapi antarmuka berbasis web yang disebut biblioshiny. Aplikasi ini mudah digunakan dan memungkinkan individu tanpa pengetahuan pemrograman untuk melakukan analisis bibliometrik dengan mudah. Dengan memanfaatkan antarmuka biblioshiny, pengguna dapat mengimpor data yang berasal dari basis data Scopus dalam format *comma separated values* (CSV).

HASIL DAN PEMBAHASAN

Karakteristik Publikasi Ilmiah DIF

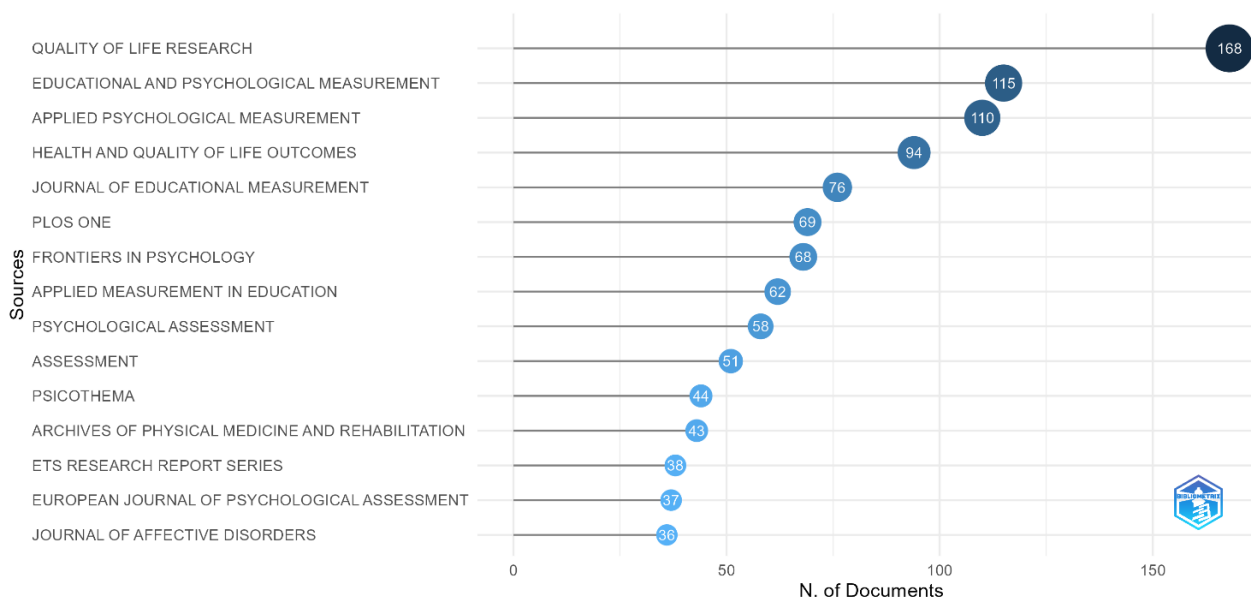
Tabel 1 menunjukkan bahwa terdapat 4.056 publikasi ilmiah yang menggunakan kata kunci *differential item functioning* dalam basis data Scopus. Publikasi ilmiah tersebut berasal dari 1.172 penerbit, 12.354 penulis, 6.304 kata kunci pilihan penulis, dan 137.297 referensi. Hal ini mencerminkan minat publikasi ilmiah yang tinggi terhadap topik DIF selama kurun waktu 1986 hingga pertengahan 2024.

Tabel 1. Karakteristik Data Publikasi Ilmiah DIF

Deskripsi	Hasil
MAIN INFORMATION ABOUT DATA	
<i>Timespan</i>	1986:2024
<i>Sources (Journals, Books, etc.)</i>	1.172
<i>Documents</i>	4.056
<i>Annual Growth Rate %</i>	14,63
<i>Document Average Age</i>	8,67
<i>Average citations per doc</i>	21,75
<i>References</i>	137.297
DOCUMENT CONTENTS	
<i>Keywords Plus (ID)</i>	6.768
<i>Author's Keywords (DE)</i>	6.304
AUTHORS	
<i>Authors</i>	12.354
<i>Authors of single-authored docs</i>	250
AUTHORS COLLABORATION	
<i>Single-authored docs</i>	345
<i>Co-Authors per Doc</i>	4,72
<i>International co-authorships %</i>	30,05
DOCUMENT TYPES	
<i>Article</i>	3.979
<i>Conference Paper</i>	77

Gambar 2 menunjukkan bahwa *Quality of Life Research* (168 publikasi), *Educational and Psychological Measurement* (115 publikasi), dan *Applied Psychological Measurement* (110 publikasi) adalah tiga penerbit yang paling banyak memuat publikasi ilmiah DIF. Jurnal lainnya adalah *Health and Quality of Life Outcomes* (94 publikasi), *Journal of Educational Measurement* (76 publikasi), *PLoS ONE* (69 publikasi), *Frontiers in Psychology* (68 publikasi), *Applied Measurement in Education* (62 publikasi), *Psychological*

Assessment (58 publikasi), dan *Assessment* (51 publikasi). Saluran publikasi ini dapat menjadi referensi utama bagi penulis yang ingin menerbitkan penelitian yang berkaitan dengan DIF.



Gambar 2. Penerbit yang Paling Relevan

Tabel 2 menunjukkan indeks-h masing-masing publikasi yang paling relevan. Hirsch (2005) mengusulkan pendekatan baru untuk mengukur kinerja penelitian individu di tingkat mikro. Pengukuran baru ini, yang dikenal sebagai Indeks Hirsch atau indeks-h, mempertimbangkan visibilitas dan kuantitas karya seorang penulis. Indeks-h menunjukkan bahwa seorang penulis telah menulis setidaknya satu artikel yang dikutip setidaknya satu kali. *Quality of Life Research* adalah penerbit yang paling berpengaruh, dengan 4.378 kutipan dan h-indeks 34. Artinya, 34 artikel yang diterbitkan di jurnal tersebut setidaknya telah dikutip lima kali. *Quality of Life Research* mengungguli *Journal of Educational Measurement*, *Applied Measurement in Education*, *Applied Psychological Measurement*, dan *Educational and Psychological Measurement* dalam perolehan indeks-h meskipun jurnal-jurnal tersebut telah menerbitkan publikasi ilmiah DIF sejak 1986. Hal ini mengindikasikan bahwa terjadi peningkatan minat terhadap DIF dalam penelitian kualitas hidup.

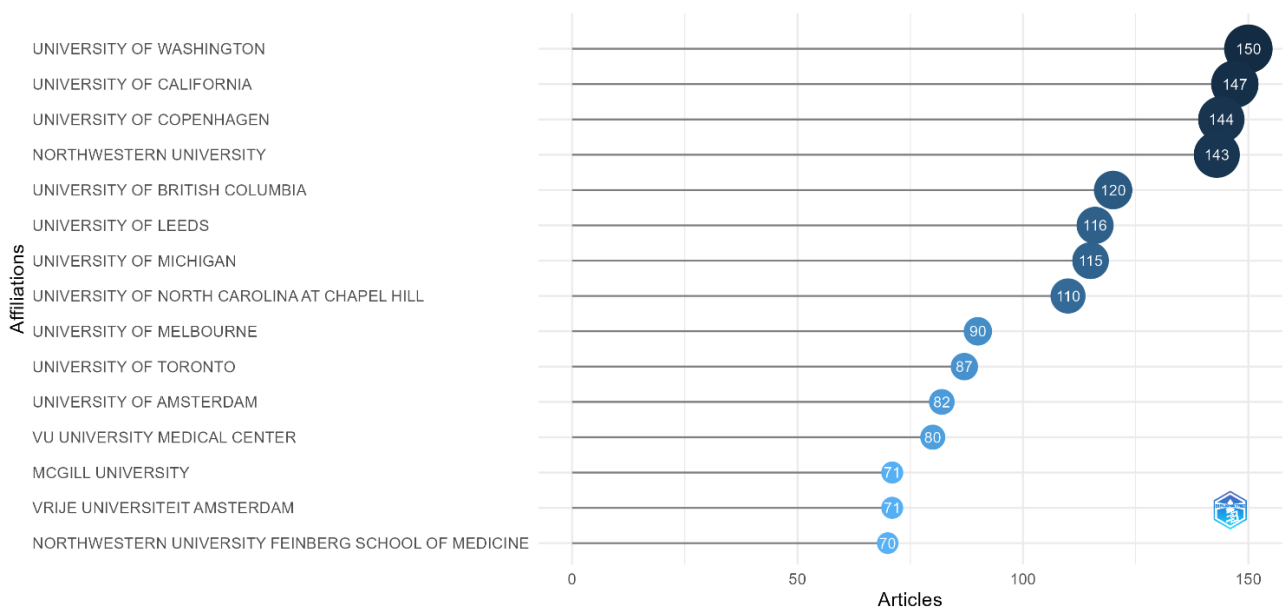
Tabel 2. Penerbit yang Paling Memberikan Dampak

Penerbit	h index	g index	m index	TC	NP	PY start
<i>Quality of Life Research</i>	34	58	1,42	4.378	168	2001
<i>Applied Psychological Measurement</i>	31	56	0,89	3.525	110	1990
<i>Journal of Educational Measurement</i>	29	56	0,74	3.240	76	1986
<i>Health and Quality of Life Outcomes</i>	26	42	1,3	2.141	94	2005
<i>Applied Measurement in Education</i>	25	41	0,69	1.798	62	1989
<i>Educational and Psychological Measurement</i>	25	38	0,74	2.093	115	1991
<i>Psychological Assessment</i>	21	50	0,78	2.589	58	1998

<i>Archives of Physical Medicine and Rehabilitation</i>	20	32	0,95	1.07 5	43	2004
<i>Journal of Clinical Epidemiology</i>	20	35	0,67	1.79 4	35	1995
<i>PLoS ONE</i>	19	36	1,46	1.41 3	69	2012

Catatan. TC = total citations. NP = number of publications. PY = year of starting publication.

Gambar 3 menunjukkan 10 institusi teratas yang memiliki peran penting dalam publikasi ilmiah DIF. Terlihat bahwa University of Washington (Amerika Serikat) berada di posisi pertama dengan 150 publikasi, diikuti University of California (Amerika Serikat) dengan 147 publikasi, University of Copenhagen (Denmark) dengan 144 publikasi, Northwestern University (Amerika Serikat) dengan 143 publikasi, University of British Columbia (Amerika Serikat) dengan 120, University of Leeds (Inggris) dengan 116 publikasi, University of Michigan (Amerika Serikat) dengan 115 publikasi, University of North Carolina at Chapel Hill (Amerika Serikat) dengan 110 publikasi, University of Melbourne (Australia) dengan 90 Publikasi, dan dan yang terakhir adalah University of Toronto (Kanada) dengan 87 publikasi. Institusi-institusi tersebut sebagian besar berada di Amerika Serikat, sehingga dapat disimpulkan bahwa publikasi ilmiah DIF relatif lebih berkembang dan menjadi perhatian di Amerika Serikat.



Gambar 3. Afiliasi yang Paling Relevan

Dalam beberapa dekade terakhir, kolaborasi ilmiah internasional telah berkembang pesat di seluruh dunia (Witze, 2016). Kolaborasi ilmiah ditandai dengan adanya dua atau lebih afiliasi dan penulis dalam suatu publikasi ilmiah. Publikasi dengan kolaborasi ilmiah internasional mempunyai pengaruh yang lebih signifikan dibandingkan dengan publikasi dengan kolaborasi nasional (Barrantes dkk., 2012). Gambar 4 menyajikan kolaborasi publikasi ilmiah DIF di berbagai negara.



Gambar 4. Peta Kolaborasi Antar negara

Gambar 4 memvisualisasikan negara yang berkolaborasi secara internasional dengan area berwarna biru. Sementara itu, pita berwarna cokelat menghubungkan negara yang mewakili tingkat kerja sama di antara para penulis yang terlibat dalam suatu penelitian. Amerika Serikat-Kanada memiliki jumlah kolaborasi yang paling banyak (99), diikuti Amerika Serikat-Inggris (68), Amerika Serikat-Belanda (61), Inggris-Australia (57), Amerika Serikat-Cina (55), Amerika Serikat-Australia (48), Amerika Serikat-Jerman (47), Inggris-Swedia (43), Amerika Serikat-Swedia (43), dan Inggris-Jerman (41). Para penulis dari Amerika Serikat bersama rekan-rekannya memiliki kontribusi yang signifikan dalam perkembangan topik penelitian DIF secara global.

Tabel 3 menyajikan kontribusi yang dibuat oleh penulis dari suatu negara. Terlihat bahwa Amerika Serikat memiliki jumlah penulis sebagai koresponden terbanyak dengan total 1.167 publikasi. Secara khusus, 980 publikasi diterbitkan penulis dalam negeri dan 187 publikasi diterbitkan secara kolektif dengan penulis yang berasal dari berbagai negara. Hal ini diikuti dengan Inggris (93 SCP dan 84 MCP), Cina (117 SCP dan 55 MCP), Australia (101 SCP dan 65 MCP), dan Kanada (87 SCP dan 75 MCP).

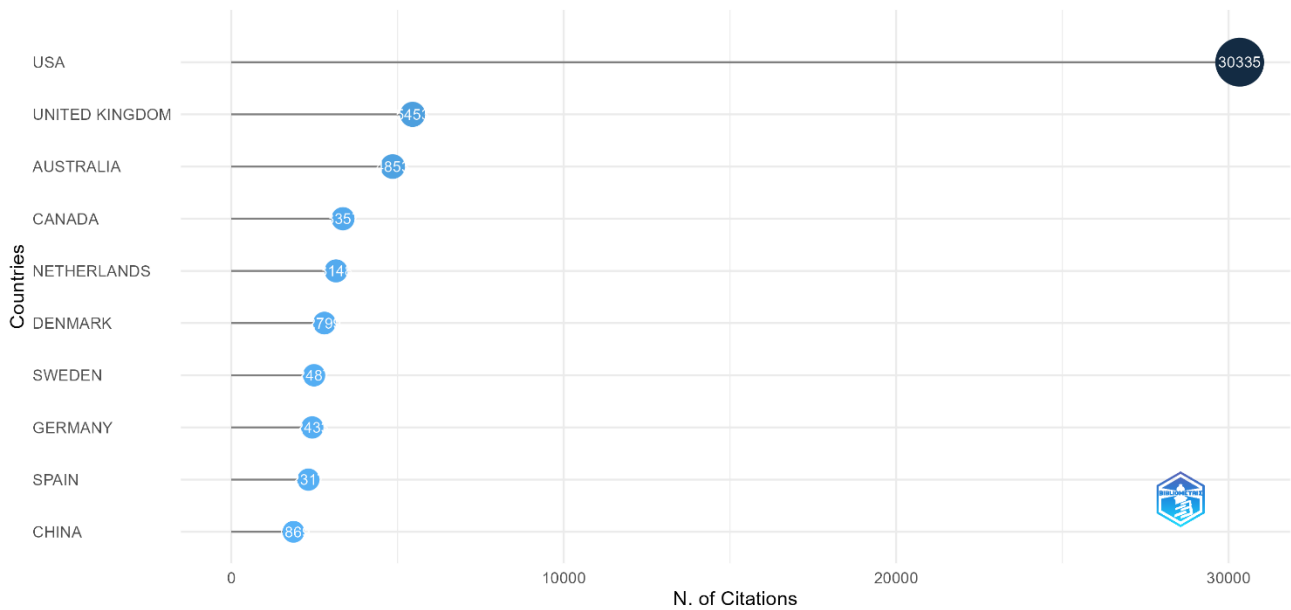
Tabel 3. Negara yang Paling Kolaboratif

Negara	Publikasi	SCP	MCP	Frekuensi	Rasio MCP
Amerika Serikat	1.167	980	187	0,29	0,16
Inggris	177	93	84	0,04	0,48
Cina	172	117	55	0,04	0,32
Australia	166	101	65	0,04	0,39
Kanada	162	87	75	0,04	0,46
Spanyol	159	114	45	0,04	0,28
Belanda	143	90	53	0,04	0,37

Jerman	131	78	53	0,03	0,41
Swedia	91	43	48	0,02	0,53
Denmark	89	63	26	0,02	0,29

Catatan. SCP = single country publications. MCP = multi country publications.

Jumlah kutipan publikasi ilmiah bergantung pada topik penelitian, bidang penelitian, dan disiplin ilmu penelitian (Dorta-González dkk., 2014). Gambar 5 menunjukkan negara-negara yang paling banyak dikutip. Amerika Serikat berada di posisi teratas dengan 30.335 kutipan, disusul Inggris 5.453, Australia 4.853 kutipan, Kanada 3.357 kutipan, Belanda 3.148 kutipan, Denmark 2.799 kutipan, Swedia 2.487 kutipan, Jerman 2.433, Spanyol 2.317, dan Cina 1.869 kutipan. Hasil ini menunjukkan bahwa Amerika Serikat memegang posisi menonjol dalam hal publikasi ilmiah DIF dengan jumlah pengutipan tertinggi.



Gambar 5. Negara yang Paling Banyak Dikutip

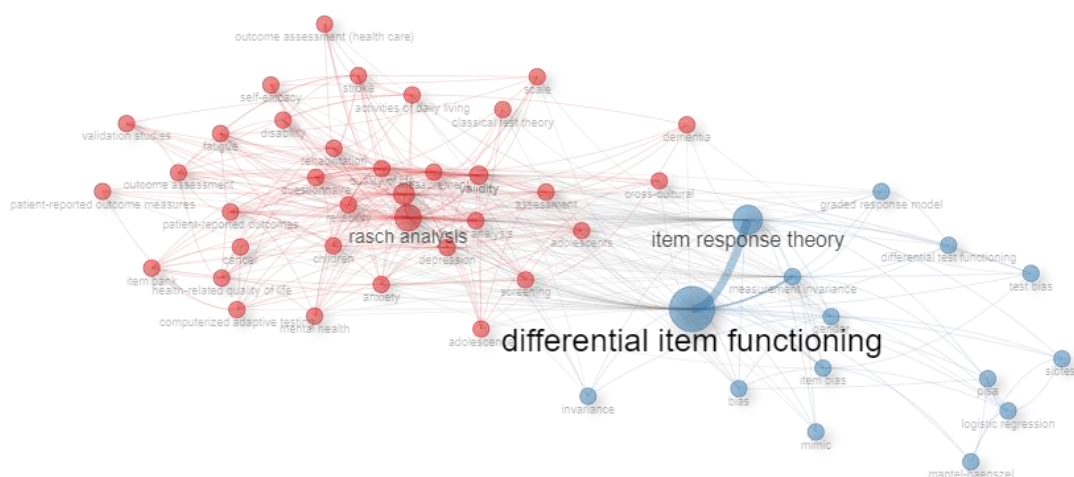
Tabel 4 menunjukkan publikasi ilmiah DIF yang paling banyak dikutip hingga pertengahan tahun 2024. Pertama adalah penelitian De Morton (2009), kemudian Reeve, Hays, Bjorner, Cook, Crane, Teresi, Thissen, Revicki, Weiss, Hambleton, Liu, Gershon, Reise, Lai, Cella, dan PROMIS Cooperative Group (2007), Pejtersen, Kristensen, Borg, dan Bjorne (2010), Swaminathan dan Rogers (1990), Marsh, Lüdtke, Muthén, Asparouhov, Morin, Trautwein, dan Nagengast (2010), Hays, Morales, dan Reise (2000), Kahler, Strong, dan Read (2005), Huang, Chung, Kroenke, Delucchi, dan Spitzer (2006), Choi, Gibbons, dan Crane (2011), dan Eremenco, Cella, dan Arnold (2005). Terlihat bahwa ada peningkatan minat penelitian DIF dalam bidang kesehatan yang ditandai dengan banyaknya kutipan pada artikel “*The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic Study*” dan “*Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS)*.”

Tabel 4. Publikasi Ilmiah yang Paling Banyak Dikutip

Penulis	Judul	Jumlah Kutipan (JK)	JK Per Tahun	JK Ternormalisasi
De Morton (2009)	<i>“The PEDro scale is a valid measure of the</i>	1.401	87,56	31,58

	<i>methodological quality of clinical trials: a demographic study</i>			
Reeve, Hays, Bjorner, Cook, Crane, Teresi, Thissen, Revicki, Weiss, Hambleton, Liu, Gershon, Reise, Lai, Cella, dan PROMIS Cooperative Group (2007)	<i>“Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS)”</i>	1.237	68,72	22,13
Pejtersen, Kristensen, Borg, dan Bjorne (2010)	<i>“The second version of the Copenhagen Psychosocial Questionnaire”</i>	871	58,07	21,00
Swaminathan dan Rogers (1990)	<i>“Detecting differential item functioning using logistic regression procedures”</i>	734	20,97	5,69
Marsh, Lüdtke, Muthén, Asparouhov, Morin, Trautwein, dan Nagengast (2010)	<i>“A new look at the big five factor structure through exploratory structural equation modeling”</i>	691	46,07	16,66
Hays, Morales, dan Reise (2000)	<i>“Item response theory and health outcomes measurement in the 21st century”</i>	690	27,60	12,83
Kahler, Strong, dan Read (2005)	<i>“Toward efficient and comprehensive measurement of the alcohol problems continuum in college students: The Brief Young Adult Alcohol Consequences Questionnaire”</i>	623	31,15	10,59
Huang, Chung, Kroenke, Delucchi, dan Spitzer (2006)	<i>“Using the patient health questionnaire-9 to measure depression among racially and ethnically diverse primary care patients”</i>	542	28,53	7,71
Choi, Gibbons, dan Crane (2011)	<i>“lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations”</i>	507	36,21	13,26
Eremenco, Cella, dan Arnold (2005)	<i>“A comprehensive method for the translation and cross-cultural validation of health status questionnaires”</i>	473	23,65	8,04

Kata kunci adalah metrik, yang umum digunakan dalam basis data dan mesin pencari, yang secara konsisten memainkan peran penting dalam pengindeksan (Ghanbarpour & Naderi, 2019). Kata kunci penulis, di sisi lain, mewakili frasa yang dipilih dan dibuat secara khusus oleh penulis, berfungsi sebagai elemen penting dalam merangkum dan menyampaikan esensi suatu publikasi ilmiah (Kwon, 2018). Menganalisis kata kunci ini sangat penting karena memfasilitasi pemahaman yang cepat tentang topik dan penekanan suatu publikasi ilmiah.

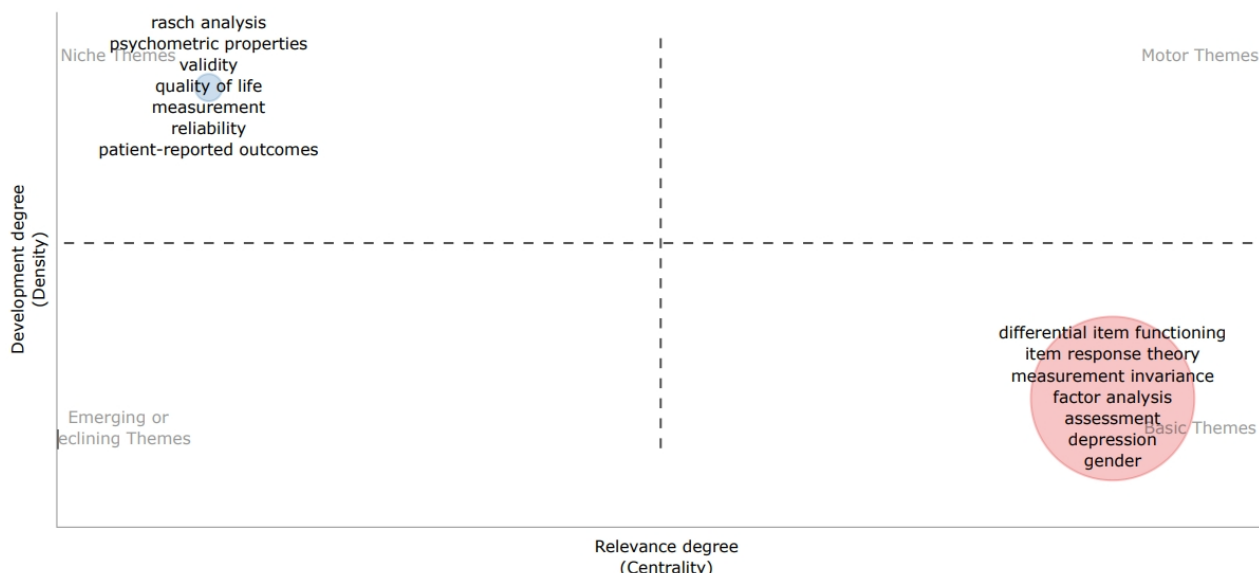


Gambar 6 adalah visualisasi jaringan kata kunci yang sering digunakan penulis dalam publikasi ilmiah DIF. Kata kunci tersebut membentuk dua klaster besar, klaster berwarna merah dan klaster berwarna biru. Klaster berwarna merah mencakup kata kunci yang berkaitan dengan pendeteksian DIF di berbagai bidang disiplin ilmu, yang meliputi *quality of life*, *depression*, *patient-reported outcomes*, dan *activities of daily living* yang mana pelaporannya berkaitan dengan dengan *rasch analysis*, *psychometric properties*, *validity*, *factor analysis* dan *assessment*. Kluster berwarna biru mencakup kata kunci yang berkaitan dengan metode pendeteksian DIF yang populer, seperti *item response theory*, *measurement invariance*, *logistic regression*, *sibtest*, dan *mantel-haenszel*.

Klaster	Kata Kunci	Tema
Merah	<i>rasch analysis, psychometric properties, validity, quality of life, factor analysis, measurement, assessment, depression, patient-reported outcomes, reliability, questionnaire, rehabilitation, anxiety, children, adolescents, health-related quality of life, mental health, screening, classical test theory, activities of daily living, stroke, disability, computerized adaptive testing, dementia, validation studies, self-efficacy, fatigue, cancer, cross-cultural, item bank, patient-reported outcome measures outcome assessment, scale, adolescence, outcome assessment (health care)</i>	Tema niche
Biru	<i>differential item functioning, item response theory, measurement invariance, gender, item bias, logistic</i>	Tema dasar

regression, sibtest, mantel-haenszel, test bias, bias, graded response model, invariance, differential test functioning, pisa, mimic

Gambar 7 adalah peta tematik berdasarkan kata kunci yang dipilih penulis yang dipetakan ke dalam empat tema besar (Alkhamash, 2023). Pertama, tema motor (kuadran pertama, kanan atas): tema yang memiliki sentralitas dan kepadatan tinggi. Hal ini menandakan tema tersebut penting dalam penelitian DIF sehingga mendorong penelitian DIF terus berkembang. Kedua, tema *niche* (kuadran kedua, kiri atas): tema dengan kepadatan tinggi dan sentralitas rendah, menandakan bahwa tema tersebut memiliki relevansi yang terbatas dalam penelitian DIF. Ketiga, tema yang muncul atau menurun (kuadran ketiga, kiri bawah): tema dengan sentralitas rendah dan kepadatan rendah, menyiratkan bahwa tema tersebut berkembang secara minimal dan marginal. Keempat, tema dasar (kuadran keempat, kanan bawah): tema tersebut memiliki sentralitas tinggi dan kepadatan rendah. Tema-tema ini sangat penting dalam memfasilitas penelitian lintas disiplin ilmu yang melibatkan DIF.



Gambar 7. Peta Tematik Penelitian DIF Berdasarkan Kata Kunci

Gambar 7 menunjukkan bahwa peta tematik hanya terisi pada kuadran kedua dan kuadran keempat. Pada kuadran kedua atau tema *niche* diwakili 1 klaster (kiri atas) yang mencakup tema-tema seperti *Rasch analysis*, *psychometric properties*, *validity*, *quality of life*, *measurement*, *reliability*, dan *patient-reported outcomes*. Sementara itu, pada kuadran keempat atau tema dasar diwakili 1 klaster (kanan bawah) yang mencakup tema-tema seperti *differential item functioning*, *item response theory*, *measurement invariance*, *factor analysis*, *assessment*, *depression*, dan *gender*. Dalam representasi visual pergerakan ke arah kanan atas dari waktu ke waktu menunjukkan tren naik, sedangkan jalur ke kiri bawah menunjukkan tren menurun.

Peta jaringan dan tematik merupakan jenis peta yang secara efektif menampilkan data yang berkaitan dengan tema atau pokok bahasan suatu penelitian. Dalam bidang bibliometrik, peta jenis ini berfungsi sebagai alat visualisasi untuk menggambarkan keterkaitan antara berbagai topik penelitian dan frekuensi kemunculannya dalam literatur ilmiah. Tingkat kemiripan antara dua publikasi dapat ditentukan dari kemunculan bersama suatu kata kunci dalam suatu publikasi. Semakin banyak kata kunci yang muncul bersamaan, semakin tinggi kemungkinan bahwa publikasi tersebut berasal dari bidang penelitian atau spesialisasi yang sama pada tingkat yang lebih tinggi (Cobo dkk., 2011).

Terkait dengan tema *niche*, hasil penelitian ini berkesesuaian dengan hasil penelitian terdahulu yang menunjukkan bahwa penelitian DIF telah berkembang dalam berbagai bidang, seperti kesehatan, epidemiologi, psikiatri dan berbagai bidang lainnya (Jones, 2013; Özmercan, 2023; Rouquette dkk., 2019; Teresi dkk., 2008). Terkait dengan pendeteksian DIF dalam bidang kesehatan, Rouquette dkk. (2019) memberikan beberapa rekomendasi dalam pendeteksian DIF: 1) Membedakan antara *uniform* DIF, *balanced non-uniform* DIF, dan *unbalanced non-uniform* DIF; 2) Ketika DIF secara statistik signifikan, maka besaran efeknya (*effect size*) perlu diperkirakan dalam kelompok dan dalam berbagai variasi parameter; 3) Ketika DIF secara statistik signifikan, maka skala yang dimodifikasi karena DIF harus dijelaskan secara rinci dan apabila tidak dimodifikasi maka perlu memberikan beberapa rekomendasi; dan 4) Jika beberapa butir soal menunjukkan DIF yang secara statistik signifikan pada dimensi yang sifatnya berlawanan, maka evaluasi perlu dilakukan pada masing-masing dimensi.

Terkait dengan tema motor, hasil penelitian ini berkesesuaian dengan hasil penelitian terdahulu yang menunjukkan bahwa *item response theory* adalah pendekatan yang paling populer dalam pendeteksian DIF (Berrío dkk., 2020; Gómez-Benito dkk., 2018; Guilera dkk., 2013; Nugent, 2017). Beberapa metode lain yang digunakan adalah *rasch analysis*, *logistic regression*, *sibtest*, dan metode Mantel-Haenszel. Jika dibandingkan dengan metode yang lain, pendekatan *item response theory* lebih fleksibel karena dapat diterapkan pada data berjenis dikotomi (*dichotomous*) maupun politomi (*polytomous*) (Berrío dkk., 2020). Menariknya, metode *rasch analysis* menjadi salah satu kata kunci yang muncul di tema motor. Hal ini menunjukkan bahwa metode *rasch analysis* telah menjadi metode yang populer dalam pendeteksian DIF di beberapa disiplin ilmu, khususnya kualitas hidup dan laporan hasil kesehatan pasien.

Keterbatasan penelitian ini terletak pada penggunaan jenis publikasi berupa artikel jurnal dan prosiding konferensi. Sehingga, hasil penelitian ini mungkin tidak merepresentasikan kondisi publikasi ilmiah DIF secara menyeluruh. Oleh karena itu, penelitian selanjutnya dapat melakukan tinjauan kritis publikasi ilmiah DIF menggunakan basis data selain basis data populer (Scopus dan WoS), seperti Google Scholar dan Microsoft Academic yang memuat jenis publikasi ilmiah lebih beragam (laporan penelitian yang tidak diterbitkan, tesis, dan disertasi mahasiswa). Penelitian yang dilakukan Visser dkk. (2021) menunjukkan bahwa Microsoft Academic sejauh ini menawarkan cakupan literatur ilmiah yang paling komprehensif dibandingkan Scopus, WoS, Dimensions, dan Crossref. Meskipun demikian, penggunaan jenis publikasi ilmiah berupa jurnal dan prosiding konferensi meningkatkan kredibilitas gambaran perkembangan publikasi ilmiah DIF dari tahun 1986 hingga pertengahan 2024. Penelitian yang akan datang juga dapat membandingkan hasil penelitian ini dengan perkembangan publikasi ilmiah DIF beberapa dekade ke depan.

SIMPULAN

Hasil analisis bibliometrik menunjukkan bahwa Amerika Serikat memiliki peran sentral dalam perkembangan publikasi ilmiah DIF. Hal ini terlihat dari jumlah publikasi, jumlah afiliasi, dan jumlah kutipan yang didapatkannya. Amerika Serikat juga menduduki posisi tertinggi dalam peta kolaborasi antarnegara. Berdasarkan analisis jaringan kemunculan kata kunci dan peta tematik, terdapat dua tema utama dalam publikasi ilmiah DIF. Pertama, tema *niche* yang mencakup tema-tema seperti *Rasch analysis*, *psychometric properties*, *validity*, *quality of life*, *measurement*, *reliability*, dan *patient-reported outcomes*. Kedua, tema dasar yang mencakup tema-tema seperti *differential item functioning*, *item response theory*, *measurement invariance*, *factor analysis*, *assessment*, *depression*, dan *gender*. Hasil ulasan ini dapat menjadi rujukan para peneliti yang tertarik dengan penelitian DIF.

DAFTAR PUSTAKA

Alkhamash, R. (2023). Bibliometric, Network, And Thematic Mapping Analyses Of Metaphor And Discourse In Covid-19 Publications From 2020 To 2022. *Frontiers In Psychology*, 13, 1062943. <https://doi.org/10.3389/fpsyg.2022.1062943>

- American Educational Research Association, American Psychological Association, & National Council On Measurement In Education. (2014). *Standards For Educational And Psychological Testing*. American Educational Research Association.
- Angoff, W. H. (1993). Perspectives On Differential Item Functioning Methodology. Dalam P. W. Holland & Howard Wainer (Ed.), *Differential Item Functioning*. (Hlm. 3–23). Lawrence Erlbaum Associates.
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-Tool For Comprehensive Science Mapping Analysis. *Journal Of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/J.Joi.2017.08.007>
- Berrío, Á. I., Gómez-Benito, J., & Arias-Patiño, E. M. (2020). Developments And Trends In Research On Methods Of Detecting Differential Item Functioning. *Educational Research Review*, 31, 100340. <https://doi.org/10.1016/J.Edurev.2020.100340>
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R Package For Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory And Monte Carlo Dimulations. *Journal Of Statistical Software*, 39(8). <https://doi.org/10.18637/Jss.V039.I08>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An Approach For Detecting, Quantifying, And Visualizing The Evolution Of A Research Field: A Practical Application To The Fuzzy Sets Theory Field. *Journal Of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/J.Joi.2010.10.002>
- De Morton, N. A. (2009). The Pedro Scale Is A Valid Measure Of The Methodological Quality Of Clinical Trials: A Demographic Study. *Australian Journal Of Physiotherapy*, 55(2), 129–133. [https://doi.org/10.1016/S0004-9514\(09\)70043-1](https://doi.org/10.1016/S0004-9514(09)70043-1)
- Dorta-González, P., Dorta-González, M. I., Santos-Peñate, D. R., & Suárez-Vega, R. (2014). Journal Topic Citation Potential And Between-Field Comparisons: The Topic Normalized Impact Factor. *Journal Of Informetrics*, 8(2), 406–418. <https://doi.org/10.1016/J.Joi.2014.01.013>
- Eremenco, S. L., Cella, D., & Arnold, B. J. (2005). A Comprehensive Method For The Translation And Cross-Cultural Validation Of Health Status Questionnaires. *Evaluation & The Health Professions*, 28(2), 212–232. <https://doi.org/10.1177/0163278705275342>
- Ghanbarpour, A., & Naderi, H. (2019). A Model-Based Method To Improve The Quality Of Ranking In Keyword Search Systems Using Pseudo-Relevance Feedback. *Journal Of Information Science*, 45(4), 473–487. <https://doi.org/10.1177/0165551518799637>
- Gómez-Benito, J., Sireci, S., Padilla, J.-L., Hidalgo, M. D., & Benítez, I. (2018). Differential Item Functioning: Beyond Validity Evidence Based On Internal Structure. *Psicothema*, 30(1), 104–109.
- Guilera, G., Gómez-Benito, J., Hidalgo, M. D., & Sánchez-Meca, J. (2013). Type I Error And Statistical Power Of The Mantel-Haenszel Procedure For Detecting Dif: A Meta-Analysis. *Psychological Methods*, 18(4), 553–571. <https://doi.org/10.1037/A0034306>
- Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item Response Theory And Health Outcomes Measurement In The 21st Century. *Medical Care*, 38, ii-28-ii-42. <https://doi.org/10.1097/00005650-200009002-00007>
- Hirsch, J. E. (2005). An Index To Quantify An Individual's Scientific Research Output. *Proceedings Of The National Academy Of Sciences*, 102(46), 16569–16572. <https://doi.org/10.1073/Pnas.0507655102>
- Huang, F. Y., Chung, H., Kroenke, K., Delucchi, K. L., & Spitzer, R. L. (2006). Using The Patient Health Questionnaire-9 To Measure Depression Among Racially And Ethnically Diverse Primary Care Patients. *Journal Of General Internal Medicine*, 21(6), 547–552. <https://doi.org/10.1111/J.1525-1497.2006.00409.X>
- Jones, R. A. (2013). Jung's "Psychology With The Psyche" And The Behavioral Sciences. *Behavioral Sciences*, 3(3), 408–417. <https://doi.org/10.3390/Bs3030408>

- Kahler, C. W., Strong, D. R., & Read, J. P. (2005). Toward Efficient And Comprehensive Measurement Of The Alcohol Problems Continuum In College Students: The Brief Young Adult Alcohol Consequences Questionnaire. *Alcoholism: Clinical And Experimental Research*, 29(7), 1180–1189. <https://doi.org/10.1097/01.Alc.0000171940.95813.A5>
- Kwon, S. (2018). Characteristics Of Interdisciplinary Research In Author Keywords Appearing In Korean Journals. *Malaysian Journal Of Library And Information Science*, 23(2), Article 2. <https://doi.org/10.22452/Mjlis.Vol23no2.5>
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A New Look At The Big Five Factor Structure Through Exploratory Structural Equation Modeling. *Psychological Assessment*, 22(3), 471–491. <https://doi.org/10.1037/A0019227>
- Mumpuni, K. E., Begimbetova, G. A., & Retnawati, H. (2023). Self-Regulated Learning Questionnaire: Differential Item Functioning (Dif) And Calibration Using Rasch Model Analysis. *Jurnal Varidika*, 35(1), 65–80. <https://doi.org/10.23917/Varidika.V1i1.22995>
- Nugent, W. R. (2017). Understanding Dif And Dtf: Description, Methods, And Implications For Social Work Research. *Journal Of The Society For Social Work And Research*, 8(2), 305–334. <https://doi.org/10.1086/691525>
- Özmercan, E. E. (2023). The Bibliometric Analysis Of The Differential Item Functioning Using Vosviewer. *International Journal Of Eurasian Education And Culture*, 8(23), 2172–2190. <https://doi.org/10.35826/Ijoec.794>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The Prisma 2020 Statement: An Updated Guideline For Reporting Systematic Reviews. *Bmj*, 372, N71. <https://doi.org/10.1136/Bmj.N71>
- Pejtersen, J. H., Kristensen, T. S., Borg, V., & Bjorner, J. B. (2010). The Second Version Of The Copenhagen Psychosocial Questionnaire. *Scandinavian Journal Of Public Health*, 38(3_Suppl), 8–24. <https://doi.org/10.1177/1403494809349858>
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., Thissen, D., Revicki, D. A., Weiss, D. J., Hambleton, R. K., Liu, H., Gershon, R., Reise, S. P., Lai, J., & Cella, D. (2007). Psychometric Evaluation And Calibration Of Health-Related Quality Of Life Item Banks: Plans For The Patient-Reported Outcomes Measurement Information System (Promis). *Medical Care*, 45(5), S22–S31. <https://doi.org/10.1097/01.Mlr.0000250483.85507.04>
- Rouquette, A., Hardouin, J.-B., Vanhaesebrouck, A., Sébille, V., & Coste, J. (2019). Differential Item Functioning (Dif) In Composite Health Measurement Scale: Recommendations For Characterizing Dif With Meaningful Consequences Within The Rasch Model Framework. *Plos One*, 14(4), E0215073. <https://doi.org/10.1371/Journal.Pone.0215073>
- Sari, A. A., & Rahmawati, R. (2023). A Bibliometric Analysis Of Aloe Vera In Wound Healing. *Pharmacon: Jurnal Farmasi Indonesia*, 20(1), 36–49. <https://doi.org/10.23917/Pharmacon.V20i1.22312>
- Supriani, I., Iswati, S., Bella, F. I., & Tumewang, Y. K. (2022). A Bibliometric Analysis Of Zakat Literature From 1964 To 2021. *Journal Of Islamic Economic Laws*, 5(2), Article 2. <https://doi.org/10.23917/Jisel.V5i2.18511>
- Supriyadi, E., Inayah, S., Dahlan, J. A., & Darhim, D. (2022). Bibliometric Review: Elementary And Mathematics Education From Indonesian Authors. *Profesi Pendidikan Dasar*, 9(2), 176–190. <https://doi.org/10.23917/Ppd.V9i2.19481>
- Swaminathan, H., & Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal Of Educational Measurement*, 27(4), 361–370. <https://doi.org/10.1111/J.1745-3984.1990.Tb00754.X>

3989 *Differential Item Functioning (DIF): Sebuah Analisis Bibliometrik* - Suwanda Priyadi
DOI : <https://doi.org/10.31004/edukatif.v6i4.7463>

Teresi, J. A., Ramirez, M., Lai, J.-S., & Silver, S. (2008). Occurrences And Sources Of Differential Item Functioning (Dif) In Patient-Reported Outcome Measures: Description Of Dif Methods, And Review Of Measures Of Depression, Quality Of Life And General Health. *Psychology Science Quarterly*, 50(4), 538.

Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-Scale Comparison Of Bibliographic Data Sources: Scopus, Web Of Science, Dimensions, Crossref, And Microsoft Academic. *Quantitative Science Studies*, 2(1), 20–41. https://doi.org/10.1162/Qss_A_00112

Witze, A. (2016). Research Gets Increasingly International. *Nature*. <https://doi.org/10.1038/Nature.2016.19198>

Zupic, I., & Čater, T. (2015). Bibliometric Methods In Management And Organization. *Organizational Research Methods*, 18(3), 429–472. <https://doi.org/10.1177/1094428114562629>