

VALIDASI ASESMEN KOMPETENSI KEPERIBADIAN GURU: ANALISIS RASCH

Tb. Moh. Irma Ari Irawan¹, Hengki Satrianta¹, Hendrik Tuaputimain¹, Alice Asiedu¹, Ahman^{1*}

Program Studi Bimbingan dan Konseling, Universitas Pendidikan Indonesia¹

Email: ahman@upi.edu*

Abstrak

Penelitian ini bertujuan memvalidasi instrumen asesmen kompetensi kepribadian guru menggunakan Rasch Model. Penelitian kuantitatif ini melibatkan 165 guru melalui instrumen dengan 72 aitem untuk mengukur 13 dimensi kompetensi kepribadian. Analisis menggunakan Winsteps 5.5.0 menunjukkan *person reliability* 0.83 dan *item reliability* 0.99, mengindikasikan konsistensi pengukuran sangat baik. Hasil analisis dimensionalitas menunjukkan 66.1% varians dijelaskan oleh pengukuran dengan *unexplained variance in first contrast* 6.7%, memenuhi kriteria unidimensionalitas. Kategori respons berfungsi optimal dengan *threshold* terurut (-1.16, -0.46, 1.62). Dari 72 item, 64 item (88.9%) menunjukkan *fit* yang baik, sementara 8 item memerlukan revisi. Temuan menunjukkan dimensi “arif dan bijaksana” memiliki tingkat kesulitan tertinggi (3.42 logit), sedangkan “mengembangkan diri berkelanjutan” terendah (-2.01 logit). Instrumen ini memberikan kontribusi praktis untuk LPTK dalam mengevaluasi kompetensi kepribadian guru secara objektif dan berbasis bukti di era digital.

Kata Kunci : kompetensi kepribadian; guru; rasch model

Abstract

This study aims to validate personality competence assessment instruments for teachers with Rasch Model. This quantitative research involved 165 teachers with 72-items instrument measuring 13 dimensions of personality competence. Analysis using Winsteps 5.5.0 showed person reliability 0.83 and item reliability 0.99, indicating excellent measurement consistency. Dimensionality analysis results showed 66.1% variance explained by measurement with unexplained variance in first contrast 6.7%, meeting unidimensionality criteria. Response categories functioned optimally with ordered thresholds (-1.16, -0.46, 1.62). Of 72 items, 64 items (88.9%) showed good fit, while 8 items required revision. Findings showed “wise and prudent” dimension had highest difficulty level (3.42 logit), while “continuous professional development” lowest (-2.01 logit). This instrument provides practical contribution for teacher education institutions in evaluating teachers' personality competence objectively and evidence-based in digital era.

Key Words : personality competence; teachers; rasch model

PENDAHULUAN

Revolusi Industri 4.0 dan transisi menuju *Society 5.0* telah menghadirkan transformasi fundamental dalam ekosistem pendidikan, menuntut reorientasi kompetensi guru yang tidak hanya menguasai teknologi tetapi juga memperkuat dimensi humanistik dan kepribadian [1], [2]. Kompetensi kepribadian guru menjadi semakin krusial di era digital dimana pembelajaran tidak lagi terbatas pada ruang kelas fisik, namun meluas ke lingkungan virtual yang kompleks dan dinamis [3]. Penelitian terkini menunjukkan bahwa kompetensi

kepribadian guru berkorelasi signifikan dengan *student engagement*, *well-being* siswa, dan capaian pembelajaran abad 21 [4], [5].

Dalam konteks Indonesia, Permendikbudristek Nomor 16 Tahun 2022 tentang Standar Proses menekankan pentingnya kompetensi kepribadian guru sebagai fondasi pembelajaran yang berpusat pada peserta didik [6]. Kompetensi kepribadian didefinisikan sebagai kemampuan personal yang mencerminkan kepribadian mantap, stabil, dewasa, arif, bijaksana, berwibawa, menjadi teladan, dan

berakhlak mulia [7]. Namun demikian, studi terkini menunjukkan bahwa asesmen kompetensi kepribadian guru di Indonesia masih didominasi oleh pendekatan konvensional yang kurang mempertimbangkan properti psikometrik modern [8], [9].

Kesenjangan antara tuntutan kompetensi kepribadian guru abad 21 dan instrumen pengukuran yang tersedia menjadi problematika mendesak. Penelitian sebelumnya tentang TPACK pada guru PPG menunjukkan masih rendahnya integrasi teknologi dalam kompetensi pedagogik [10], [11]. Mayoritas penelitian menggunakan angket *self-report* yang rentan terhadap *social desirability* bias [12]. Instrumen yang secara spesifik mengukur kompetensi kepribadian guru dengan pendekatan modern seperti *Item Response Theory* (IRT) masih terbatas [13].

Rasch Model, sebagai *gold standard* dalam IRT, menawarkan solusi psikometrik superior untuk mengembangkan instrumen yang objektif, *invariant*, dan memiliki properti pengukuran interval [14], [15]. Berbeda dengan *Classical Test Theory* yang hasil pengukurannya bergantung pada sampel, Rasch Model menghasilkan estimasi parameter yang *sample-free* dan *test-free*, memungkinkan *comparability* hasil asesmen lintas waktu dan konteks [16]. Penerapan Rasch Model dalam pengembangan instrumen pendidikan telah menunjukkan efektivitas tinggi dalam mengidentifikasi kualitas item, mendeteksi bias pengukuran, dan mengoptimalkan fungsi skala rating [17].

Penelitian ini bertujuan untuk: (1) menganalisis kualitas psikometrik instrumen asesmen kompetensi kepribadian calon guru menggunakan Rasch Model; (2) mengevaluasi fungsi kategori respons instrumen; (3) mengidentifikasi item-item yang *fit* dan *misfit*; dan (4) mendeskripsikan

pola distribusi kemampuan responden dan tingkat kesulitan item.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain instrumen *development study* yang merujuk pada *Standards for Educational and Psychological Testing* [18]. Proses pengembangan instrumen dilakukan secara sistematis melalui tujuh tahapan utama, yaitu: (1) studi literatur dan analisis kebutuhan untuk mengidentifikasi konstruk teoretik kompetensi kepribadian guru; (2) penyusunan *blueprint* instrumen; (3) penulisan butir; (4) *expert judgment* oleh para ahli bimbingan dan psikometri; (5) *cognitive interviews* dengan 15 responden untuk memastikan *clarity* dan *interpretability*; (6) *pilot testing* pada 50 responden; dan (7) *field testing* dengan analisis Rasch Model [14].

Partisipan penelitian terdiri atas 165 guru dari tiga universitas di wilayah Jawa, yaitu 55 dari Jawa Barat, 55 dari Jawa Tengah, dan 55 dari Jawa Timur yang dipilih menggunakan *purposive sampling*. Secara demografis, partisipan terdiri dari 68% perempuan dan 32% laki-laki dengan rata-rata usia 21,3 tahun ($SD = 1,2$). Jumlah sampel memenuhi minimum *guidelines* untuk analisis Rasch [19].

Instrumen yang dikembangkan, yaitu *Teacher Personality Competence Scale* (TPCS), terdiri dari 72 butir yang mengukur 13 dimensi: (1) beriman dan bertakwa; (2) berakhlak mulia; (3) arif dan bijaksana; (4) demokratis; (5) mantap; (6) berwibawa; (7) stabil; (8) dewasa; (9) jujur; (10) sportif; (11) teladan; (12) objektif dalam evaluasi diri; dan (13) *continuous professional development*. Setiap dimensi terdiri dari 4–6 butir, tersusun secara seimbang antara *favorable* dan *unfavorable* items. Instrumen menggunakan skala Likert 4 kategori (1 = Sangat Tidak Setuju hingga 4 = Sangat

Setuju) untuk meminimalkan *central tendency* bias [20].

Pengumpulan data dilakukan secara daring melalui *google forms* dengan beberapa pengamanan kualitas data, yaitu: mewajibkan setiap respon (*mandatory response*), pengacakan urutan butir, penyisipan *attention check items*, serta penggunaan *time stamp* untuk mendeteksi respons tidak wajar [21]. Izin etik diperoleh dari komite etik universitas, dan seluruh partisipan memberikan *digital informed consent*.

Analisis data dilakukan menggunakan *Winsteps 5.5.0* [22]. Tahapan analisis meliputi: (1) pemeriksaan kualitas data untuk mengidentifikasi *aberrant responding*; (2) kalibrasi model Rasch menggunakan *Joint Maximum Likelihood Estimation*; (3) analisis reliabilitas dan *separation*; (4) pengujian unidimensionalitas melalui *Principal Component Analysis of Residuals* (PCAR); (5) analisis item fit menggunakan kriteria *Infit/Outfit MNSQ* 0,5–1,5; (6) diagnostik fungsi kategori respons; (7) analisis *Differential Item Functioning* (DIF) berdasarkan *gender* dan *region*; serta (8) interpretasi *Wright Map* untuk menilai *person-item targeting* [23].

HASIL DAN PEMBAHASAN

Analisis reliabilitas menunjukkan hasil *excellent* dengan *person separation reliability* 0.83 dan *item separation reliability* 0.99. *Person reliability* 0.83 mengindikasikan konsistensi internal yang baik (bagus) dalam membedakan tingkat kompetensi kepribadian antarresponden, melampaui batas minimal 0.80 untuk instrumen penelitian pendidikan [19]. *Item reliability* 0.99 menunjukkan konsistensi sangat baik dalam hierarki tingkat kesulitan aitem, mengindikasikan urutan tingkat kesulitan item akan tetap konsisten jika

diujikan pada sampel berbeda dengan karakteristik serupa [24].

Person separation sebesar 2.24 menunjukkan bahwa instrumen mampu membedakan responden ke dalam sekitar 3.3 $[(4 \times 2.24 + 1) / 3 = 3.3]$ strata tingkat kemampuan [25]. Hal ini mengindikasikan bahwa instrumen dapat mengklasifikasikan calon guru ke dalam kategori kompetensi kepribadian rendah, sedang, dan tinggi secara reliabel. *Person separation* yang lebih tinggi akan lebih baik untuk tujuan *high-stakes assessment*, namun nilai 2.24 memadai untuk tujuan penelitian dan pengembangan [26].

Item separation sebesar 12.61 menunjukkan bahwa instrumen memiliki rentang tingkat kesulitan item yang sangat luas, mampu membedakan item ke dalam sekitar 17 strata tingkat kesulitan $[(4 \times 12.61 + 1) / 3 = 17.1]$ [25]. *Item separation* yang sangat tinggi ini mengindikasikan bahwa item-item dalam instrumen memiliki variasi tingkat kesulitan yang sangat baik, dari item yang sangat mudah hingga sangat sulit, sehingga dapat mengukur responden dengan berbagai tingkat kemampuan secara akurat. Selain itu, nilai Cronbach's Alpha yang mengukur reliabilitas interaksi antara person dan item secara keseluruhan menunjukkan sebesar 0,85 menunjukkan kriteria bagus sekali memberikan bukti validitas konvergen dari perspektif *Classical Test Theory* (CTT) [27].

Pada aspek unidimensionalitas, analisis *Principal Component Analysis of Residuals* (PCAR) memberikan dukungan yang kuat terhadap asumsi unidimensionalitas instrumen. *Raw variance explained by measures* sebesar 66,1% (istimewa) yang jauh melampaui batas minimum 40%, yang mengindikasikan bahwa konstruk kompetensi kepribadian merupakan dimensi dominan dalam instrumen [17],

[28], [29]. Nilai ini bahkan lebih tinggi dibandingkan hasil, yang melaporkan *variance explained* sebesar 58,3% pada instrumen TPACK, menunjukkan kualitas unidimensionalitas instrumen yang superior [30].

Nilai *unexplained variance in the first-third contrast* sebesar 6,7%, 1,7%, dan 1,2% berada jauh di bawah ambang batas 15%, sehingga menunjukkan tidak adanya dimensi sekunder yang substantif (*absence of substantive secondary dimension*) [31]. Temuan ini selaras dengan teori yang menyatakan bahwa meskipun kompetensi kepribadian terdiri atas berbagai *facets* (13 dimensi), seluruh *facets* tersebut saling berkorelasi erat dan membentuk suatu *unified construct* [32].

Pada aspek *item fit statistics*, hasil analisis menunjukkan bahwa 64 butir (88,9%) memiliki *acceptable fit* terhadap model, yaitu berada dalam rentang $0,5 < \text{infit/outfit MNSQ} < 1,5$. Sementara itu, delapan butir (11,1%) teridentifikasi mengalami *misfit* (lihat Tabel 1). Proporsi *misfit* ini sebanding dengan temuan Azizah et al. yang melaporkan tingkat *misfit* sebesar 12% pada tahap pengembangan instrumen [13]. Temuan tersebut mengindikasikan bahwa kualitas butir instrumen berada dalam batas wajar untuk fase pengembangan, dan sebagian besar butir telah menunjukkan konsistensi yang baik dengan model Rasch.

Tabel 1. Distribusi Item Berdasarkan Fit Statistics

| Kategori Fit | Jumlah Item | Keterangan |
|-----------------|-------------|---------------------------|
| <i>Good Fit</i> | 68 | $0.5 < \text{MNSQ} < 1.5$ |
| <i>Underfit</i> | 1 | $\text{MNSQ} > 1.5$ |
| <i>Overfit</i> | 2 | $\text{MNSQ} < 0.5$ |

| Kategori Fit | Jumlah Item | Keterangan |
|-----------------------|-------------|--|
| <i>PT-Measure Cor</i> | 1 | $>0.4 \text{ PT-Measure Cor} < 0.85$ Tidak bernilai negatif |

Sumber: Pengolahan data

Beberapa butir teridentifikasi memiliki *significant underfit*, yaitu P10 (*Infit* MNSQ = 1,49; *Outfit* MNSQ = 1,51). Sementara itu, butir P14 menunjukkan *extreme overfit* dengan nilai *Infit* MNSQ sebesar 0,44, yang mengindikasikan adanya *strong social desirability bias*, terutama mengingat konteks sosial Indonesia yang religius [33]. Secara keseluruhan, nilai *Mean Infit* MNSQ sebesar 1,00 (SD = 0,23) dan *Mean Outfit* MNSQ sebesar 1,01 (SD = 0,22) menunjukkan bahwa kualitas butir instrumen berada dalam kategori baik [34]. *Cognitive interview follow-up* juga mengidentifikasi bahwa butir yang mengalami *misfit* cenderung memiliki *ambiguous wording* atau mengukur *advanced competencies* yang berada di luar kemampuan umum calon guru pada tahap pendidikan awal [35].

Analisis terhadap fungsi kategori respons menunjukkan bahwa *4-point Likert scale* bekerja secara efektif. Distribusi observasi pada setiap kategori tampak seimbang, dengan seluruh kategori digunakan secara memadai, yaitu lebih dari 10% observasi pada masing-masing kategori. Kondisi ini memenuhi kriteria *good category functioning* [36]. Hal ini mengindikasikan bahwa peserta mampu membedakan pilihan respons secara konsisten, dan setiap kategori memberikan informasi yang bermakna bagi model.

Observed average measures menunjukkan progresi monotonik dari -1,97 logit pada Kategori 1 hingga 1,96 logit pada Kategori 4. Pola ini mengindikasikan bahwa responden dengan kemampuan lebih tinggi

secara konsisten memilih kategori respons yang lebih tinggi [37]. Nilai Andrich *thresholds* juga menunjukkan progresi yang terurut, yaitu $\tau_1 = -1,16$, $\tau_2 = -0,46$, dan $\tau_3 = 1,62$, yang mengonfirmasi bahwa kategori bekerja sesuai dengan intended continuum [38].

Jarak antar *adjacent thresholds* (0,70 dan 2,08 logit) memperlihatkan adanya ketidaksimetrisan dalam sebaran kategori. Jarak 0,70 logit berada sedikit di bawah rentang optimal 1,4–5,0 logit, sehingga mengisyaratkan potensi untuk melakukan category collapse antara Kategori 1 dan 2 pada pengembangan instrumen di masa depan [37]. Namun demikian, mengingat bahwa kedua kategori tersebut memiliki makna teoretis yang penting, struktur empat kategori saat ini tetap dipertahankan [20].

Wright Map memberikan gambaran komprehensif mengenai keterpaduan (*targetting*) antara kemampuan responden dan kesulitan butir. Visualisasi menunjukkan bahwa *mean person ability* berada pada 0.08 logit, sangat berdekatan dengan *mean item difficulty* sebesar 0.00 logit. Selisih 0.08 logit ini tergolong minimal sehingga dapat diinterpretasikan bahwa tingkat kesulitan instrumen telah selaras dengan profil kemampuan sampel, yang pada akhirnya memaksimalkan *measurement information* [19].

Distribusi kemampuan responden tampak membentuk pola mendekati normal, dengan nilai rata-rata 0.08 logit dan simpangan baku 0.48 logit, serta rentang kemampuan dari -1.06 hingga 2.32 logit. Sebaliknya, distribusi kesulitan item jauh lebih lebar, ditunjukkan oleh SD sebesar 1.81 logit dan rentang -3.35 logit (item termudah, P46) hingga 4.27 logit (item tersulit, P9). Rentang total 7.62 logit ini—yang lebih besar dibandingkan rentang kemampuan responden (3.38 logit)—menunjukkan bahwa instrumen memiliki *coverage* yang

baik untuk mengukur peserta dengan tingkat kemampuan yang sangat bervariasi [23].

Meskipun demikian, inspeksi lebih lanjut terhadap peta *person-item* mengungkap adanya *gaps* pada area tertentu, khususnya pada rentang -2.0 hingga -1.0 logit dan 2.5 hingga 4.0 logit, di mana kepadatan item terlihat rendah [39]. Keberadaan *gaps* ini berpotensi menurunkan *measurement precision* karena kemampuan responden pada rentang tersebut tidak dilayani oleh butir dengan tingkat kesulitan yang memadai. Oleh sebab itu, pengembangan butir tambahan pada rentang tersebut direkomendasikan untuk meningkatkan cakupan pengukuran sekaligus memperhalus item *targetting* agar instrumen semakin optimal dalam mengukur kompetensi kepribadian calon guru di era Revolusi Industri 4.0.

Analisis tingkat kesulitan butir berdasarkan dimensi kompetensi memberikan gambaran penting mengenai *developmental priorities* dalam penguatan kompetensi kepribadian guru. Dimensi “Arif dan Bijaksana” muncul sebagai dimensi dengan tingkat kesulitan paling tinggi, dengan *mean item difficulty* sebesar 3.42 logit dan SD 0.54. Temuan ini menunjukkan bahwa kompetensi yang berkaitan dengan kebijaksanaan merupakan area yang paling menantang untuk dikuasai oleh calon guru.

Tingginya tingkat kesulitan pada dimensi ini sejalan dengan literatur bahwa *wisdom-related competencies* membutuhkan pengalaman hidup yang luas, kemampuan refleksi mendalam, serta kedewasaan kognitif dan emosional. Berbagai studi menyatakan bahwa karakteristik tersebut biasanya masih dalam tahap perkembangan pada *pre-service teachers*, sehingga kompetensi kebijaksanaan relatif belum optimal [40]. Dengan demikian, hasil ini menegaskan pentingnya intervensi

pedagogis yang secara eksplisit menumbuhkan refleksi kritis, pengambilan keputusan etis, dan *situational judgment* sebagai fondasi bagi perkembangan kebijaksanaan profesional calon guru.

Dimensi “*Continuous Professional Development*” muncul sebagai dimensi yang paling mudah dikuasai, dengan *mean item difficulty* sebesar -2.01 logit. Temuan ini *reflects strong endorsement of professional development values* di kalangan calon guru. Namun, tingkat kesulitan yang relatif rendah mengindikasikan bahwa butir-butir pada dimensi ini cenderung mengukur *aspirational values* ketimbang *demonstrated behaviors* yang benar-benar tercermin dalam praktik profesional sehari-hari [41], [42]. Artinya, calon guru sangat menyetujui pentingnya pengembangan profesional berkelanjutan, tetapi persetujuan tersebut belum tentu mencerminkan kemampuan aktual dalam melakukan perencanaan, refleksi, atau pembelajaran mandiri berbasis kinerja.

Variasi tingkat kesulitan antardimensi yang mencapai 5.43 *logit spread* menunjukkan bahwa kompetensi kepribadian merupakan *complex construct* dengan beragam aspek yang membutuhkan strategi pengembangan berbeda [43]. Rentang variasi ini menegaskan bahwa beberapa kompetensi dapat dengan mudah diadopsi pada tahap awal pendidikan guru, sedangkan kompetensi lainnya memerlukan proses internalisasi yang lebih panjang, berbasis pengalaman, dan berorientasi pada pembentukan kedewasaan profesional.

Implikasinya, program pendidikan guru perlu memberikan perhatian khusus pada dimensi dengan tingkat kesulitan tinggi—khususnya kearifan dan kebijaksanaan—melalui pendekatan *case-based learning*, mentoring, dan *reflective practice*. Pendekatan ini memungkinkan calon guru

mengembangkan *situational judgment*, kemampuan mengambil keputusan etis, serta keterampilan reflektif yang menjadi fondasi bagi pembentukan kepribadian profesional yang matang. Dengan demikian, variasi kesulitan antardimensi dapat menjadi peta jalan pengembangan kompetensi kepribadian yang lebih terstruktur dan responsif terhadap kebutuhan perkembangan calon guru.

Analisis DIF (*Differential Item Functioning*) mengidentifikasi 5 butir (6.9%) yang menunjukkan *significant* DIF berdasarkan kelompok gender dan regional. Proporsi ini berada dalam kategori *acceptable range* ($<10\%$) untuk keperluan *cross-group comparison* [44], sehingga instrumen dapat dianggap cukup adil dalam mengukur kompetensi kepribadian lintas kelompok demografis.

Butir P38 yang mengukur *assertiveness* menunjukkan DIF *contrast* sebesar 0.62 logit yang mengarah pada kelompok laki-laki ($p=0.003$). Pola ini menggambarkan adanya gender *stereotypes* terkait ekspresi otoritas dan ketegasan, di mana laki-laki secara sosial lebih sering dikonstruksikan sebagai pihak yang lebih berani menyatakan pendapat atau mengambil alih kontrol [4]. Sementara itu, butir P7 yang mengukur kemampuan *incorporating local wisdom* menunjukkan DIF antarwilayah, dengan responden Jawa Timur memberikan penilaian lebih tinggi (DIF *contrast*= 0.68 , $p=0.002$). Fenomena ini berpotensi mencerminkan kuatnya penekanan pada pelestarian budaya lokal di wilayah tersebut, sehingga responden lebih mudah mengidentifikasi dan menyetujui pernyataan terkait nilai-nilai kearifan lokal [33].

Secara keseluruhan, tingkat kejadian DIF sebesar 6,9% mengindikasikan bahwa instrumen ini *relatively fair* untuk digunakan pada responden dengan latar

belakang demografis berbeda [45]. Meski demikian, butir-butir yang teridentifikasi mengalami DIF tetap perlu diberi *content review* untuk menentukan apakah perbedaan tersebut mencerminkan variasi yang relevan dengan konstruk atau justru merupakan *construct-irrelevant bias* [46]. Proses ini penting untuk memastikan bahwa instrumen benar-benar mengukur kompetensi kepribadian, bukan dipengaruhi oleh norma sosial, konteks budaya, atau stereotip tertentu.

Temuan bahwa *wisdom-related competencies* merupakan dimensi yang secara substansial lebih sulit memiliki implikasi penting bagi *teacher education curriculum design*. Program pendidikan guru perlu menyediakan kesempatan belajar yang intens dan berkelanjutan untuk mengembangkan kebijaksanaan praktis melalui beberapa strategi. Pertama, penerapan *case-based learning* yang melibatkan *complex ethical dilemmas* dapat membantu calon guru mengembangkan kemampuan mempertimbangkan berbagai perspektif sebelum mengambil keputusan [40]. Kedua, program mentoring oleh guru berpengalaman memungkinkan terjadinya transfer nilai, pola pikir, dan prinsip kebijaksanaan yang tidak dapat dipelajari hanya melalui teori [47]. Ketiga, penerapan *reflective practice protocols* membantu guru merefleksikan tindakan, emosi, dan keputusan mereka secara sistematis. Keempat, *supervised field experiences* memberikan konteks autentik untuk mempraktikkan kebijaksanaan dalam interaksi nyata dengan siswa, orang tua, dan komunitas sekolah [48].

Sementara itu, rendahnya tingkat kesulitan pada dimensi *continuous professional development* menunjukkan perlunya mendorong mahasiswa untuk menerjemahkan nilai-nilai pengembangan diri menjadi perilaku profesional yang berkelanjutan. Hal ini dapat dilakukan

dengan mengembangkan *specific professional learning plans* yang mengarahkan calon guru pada tujuan belajar jangka panjang [41]. Program juga perlu membangun *communities of practice* [49] yang memungkinkan mahasiswa belajar bersama, saling berbagi pengalaman, dan memecahkan masalah secara kolaboratif. Selain itu, dosen perlu *modeling teacher educator engagement* [50] agar mahasiswa melihat langsung bagaimana profesional pembelajaran terus memperbarui pengetahuan dan kompetensinya.

Perbedaan gender yang teridentifikasi dalam analisis DIF mengindikasikan perlunya menangani dampak gender *socialization* dalam pembentukan kompetensi calon guru. Program pendidikan guru perlu secara eksplisit membahas gender *stereotypes* yang dapat memengaruhi cara mahasiswa menilai dirinya maupun menjalankan peran profesional. Selain itu, perlu disediakan *safe spaces* untuk mengembangkan kompetensi yang selama ini dianggap tidak lazim bagi gender tertentu, serta menghadirkan *diverse role models* yang menggambarkan bahwa semua kompetensi kepribadian dapat dikembangkan oleh siapa pun tanpa memandang gender [51]. Dengan demikian, kurikulum dapat lebih responsif terhadap bias sosial sekaligus memastikan bahwa seluruh calon guru memiliki kesempatan yang setara untuk berkembang secara optimal.

SIMPULAN

Penelitian ini berhasil memvalidasi instrumen yang memiliki kekuatan psikometrik tinggi untuk menilai kompetensi kepribadian guru. Instrumen yang dikembangkan menunjukkan reliabilitas yang sangat baik, dengan nilai *person reliability* sebesar 0,83 dan *item reliability* sebesar 0,99. Selain itu, instrumen ini juga memiliki karakteristik

unidimensionalitas yang kuat, ditunjukkan oleh 66,1% varians yang ter jelaskan, serta *rating scale functioning* yang optimal. Sebanyak 88,9% butir dinyatakan memiliki *good fit*, yang mengindikasikan kualitas butir yang sangat memadai.

Temuan penelitian menunjukkan bahwa kompetensi yang berkaitan dengan *wisdom-related competencies* merupakan dimensi yang paling menantang, dengan tingkat kesulitan 3,42 logit. Sebaliknya, orientasi terhadap pengembangan profesional merupakan dimensi yang paling mudah, berada pada titik -2,01 logit. Perbedaan tingkat kesulitan antardimensi ini memberikan wawasan yang dapat ditindaklanjuti, khususnya dalam menentukan prioritas kurikulum dan area yang perlu diperkuat dalam proses pembinaan calon guru.

Instrumen ini memberikan kontribusi praktis bagi LPTK dalam melakukan evaluasi kompetensi kepribadian calon guru secara lebih objektif, terstandar, dan *evidence-based*. Profil tingkat kesulitan tiap dimensi yang dihasilkan dari analisis memberikan *roadmap* yang jelas bagi pengembangan kurikulum, membantu mengidentifikasi dimensi mana yang memerlukan pembinaan intensif dan mana yang sudah relatif berkembang dengan baik. Selain itu, penggunaan *Wright Maps* untuk memvisualisasikan distribusi *person-item* memudahkan penyampaian hasil asesmen kepada berbagai pemangku kepentingan, baik dosen, pengelola program, maupun pembuat kebijakan pendidikan.

Untuk memperkuat temuan ini, penelitian lanjutan masih diperlukan. Beberapa arah penelitian yang direkomendasikan meliputi: (1) studi longitudinal untuk melacak perkembangan kompetensi dari waktu ke waktu; (2) studi *predictive validity* untuk menguji hubungan kompetensi kepribadian dengan efektivitas mengajar; (3) penelitian

tentang efektivitas intervensi yang dirancang untuk meningkatkan kompetensi tertentu; (4) validasi lintas budaya untuk memastikan generalisasi instrumen; serta (5) pengembangan asesmen berbasis teknologi, termasuk *video-based situational judgment tests* yang dapat memberikan gambaran kompetensi dalam konteks yang lebih autentik.

DAFTAR PUSTAKA

- [1] K. Schwab, *The fourth industrial revolution*. World Economic Forum, 2016.
- [2] Y. Zhao, A. M. Pinto Llorente, and M. C. Sánchez Gómez, "Digital competence in higher education research: A systematic literature review," *Comput. Educ.*, vol. 168, p. 104212, Jul. 2021, doi: 10.1016/j.compedu.2021.104212.
- [3] S. Blömeke, G. Kaiser, J. König, and A. Jentsch, "Profiles of mathematics teachers' competence and their relation to instructional quality," *ZDM*, vol. 52, no. 2, pp. 329–342, May 2020, doi: 10.1007/s11858-020-01128-y.
- [4] R. J. Collie, A. J. Martin, J. Bobis, J. Way, and J. Anderson, "How students switch on and switch off in mathematics: exploring patterns and predictors of (dis)engagement across middle school and high school," *Educ. Psychol.*, vol. 39, no. 4, pp. 489–509, Apr. 2019, doi: 10.1080/01443410.2018.1537480.
- [5] P. A. Jennings and M. T. Greenberg, "The Prosocial Classroom: Teacher Social and Emotional Competence in Relation to Student and Classroom Outcomes," *Rev. Educ. Res.*, vol. 79, no. 1, pp. 491–525, Mar. 2009, doi: 10.3102/0034654308325693.
- [6] Kemendikbudristek, *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 16 Tahun 2022 tentang Standar Proses*. 2022.
- [7] E. Mulyasa, *Standar kompetensi dan*

- sertifikasi guru*. Bandung: PT. Remaja Rosdakarya, 2013.
- [8] R. Hidayat and Y. E. Patras, "Assessing students' mathematical problem-solving skills in the new normal: A Rasch model approach," *Pegem J. Educ. Instr.*, vol. 11, no. 4, pp. 354–364, 2021.
- [9] D. Rahmawati, A. Nuryadin, and S. S. Fadhillah, "Rasch model analysis of teacher professional competence assessment instrument in Indonesia," *Cogent Educ.*, vol. 1, p. 2166679, 10AD.
- [10] A. Akhwani and D. W. Rahayu, "Analisis Komponen TPACK Guru SD sebagai Kerangka Kompetensi Guru Profesional di Abad 21," *J. Basicedu*, vol. 5, no. 4, pp. 1918–1925, Jun. 2021, doi: 10.31004/basicedu.v5i4.1119.
- [11] S. Azhar and T. Muchtar, "Implementasi Technology, Pedagogic, and Content Knowledge (TPACK) Guru dalam Pembelajaran pada Masa Covid-19," *J. Basicedu*, vol. 6, no. 4, pp. 6932–6938, Jun. 2022, doi: 10.31004/basicedu.v6i4.3413.
- [12] A. Sulistyono, N. Mukminatien, B. Y. Cahyono, and A. Saukah, "Enhancing learners' communication skills through authentic materials," *Cogent Educ.*, vol. 9, no. 1, p. 2043596, 2022.
- [13] N. L. Azizah, F. P. Rahmawati, and S. Ridlo, "Development and validation of pedagogical competence assessment instrument using Rasch model," *J. Phys. Conf. Ser.*, vol. 1567, no. 4, p. 042025, 2020.
- [14] W. J. Boone, "Rasch Basics for the Novice," in *Rasch Measurement*, Singapore: Springer Singapore, 2020, pp. 9–30. doi: 10.1007/978-981-15-1800-3_2.
- [15] J. Metsämuuronen, "Somers' D as an Alternative for the Item–Test and Item–Rest Correlation Coefficients in the Educational Measurement Settings," *Int. J. Educ. Methodol.*, vol. 6, no. 1, pp. 207–221, Feb. 2020, doi: 10.12973/ijem.6.1.207.
- [16] V. Aryadoust, L. Y. Ng, and H. Sayama, "A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research," *Lang. Test.*, vol. 38, no. 1, pp. 6–40, Jan. 2021, doi: 10.1177/0265532220927487.
- [17] B. Sumintono and W. Widhiarso, *Aplikasi pemodelan Rasch pada assessment pendidikan*. Trim Komunikata Publishing, 2015.
- [18] A. P. Association, *Standards for Educational and Psychological Testing*. Washington, DC 20002-4242: American Psychological Association, 2014.
- [19] J. M. Linacre, "A user's guide to Winsteps Ministep Rasch-model computer programs: Program Manual 4.8.2.," winstep.com. Accessed: Oct. 23, 2025. [Online]. Available: Winsteps.com
- [20] S. Y. Y. Chyung, K. Roberts, I. Swanson, and A. Hankinson, "Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale," *Perform. Improv.*, vol. 56, no. 10, pp. 15–23, Nov. 2017, doi: 10.1002/pfi.21727.
- [21] P. G. Curran, "Methods for the detection of carelessly invalid responses in survey data," *J. Exp. Soc. Psychol.*, vol. 66, pp. 4–19, Sep. 2016, doi: 10.1016/j.jesp.2015.07.006.
- [22] J. M. Linacre, "Winsteps® Rasch measurement computer program User's Guide Version 5.5.0," winstep.com. Accessed: Oct. 23, 2025. [Online]. Available: winstep.com
- [23] T. Bond, *Applying the Rasch Model*. Routledge, 2015. doi: 10.4324/9781315814698.
- [24] B. D. Wright and M. H. Stone, *Making Measures*. The Phaneron Press, 2004.
- [25] W. P. J. Fisher, "Rating Scale Instrument Quality Criteria."

- Accessed: Oct. 23, 2025. [Online]. Available: <https://rasch.org/rmt/rmt211m.htm>
- [26] P. W. Duncan, R. K. Bode, S. Min Lai, and S. Perera, "Rasch analysis of a new stroke-specific outcome scale: the stroke impact scale 11 No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organ," *Arch. Phys. Med. Rehabil.*, vol. 84, no. 7, pp. 950–963, Jul. 2003, doi: 10.1016/S0003-9993(03)00035-2.
- [27] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *Int. J. Med. Educ.*, vol. 2, pp. 53–55, Jun. 2011, doi: 10.5116/ijme.4dfb.8dfd.
- [28] J. M. Linacre, "Sample size and item calibration (or person measure) stability," *Rasch Meas. Trans.*, vol. 7, no. 4, p. 328, 2021.
- [29] M. D. Reckase, *Multidimensional Item Response Theory*. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-89976-3.
- [30] M. Schmid, E. Brianza, and D. Petko, "Self-reported technological pedagogical content knowledge (TPACK) of pre-service teachers in relation to digital technology use in lesson plans," *Comput. Human Behav.*, vol. 115, p. 106586, Feb. 2021, doi: 10.1016/j.chb.2020.106586.
- [31] W. J. Boone and J. R. Staver, *Advances in Rasch Analyses in the Human Sciences*. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-43420-5.
- [32] F. Caena and C. Redecker, "Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (Digcompedu)," *Eur. J. Educ.*, vol. 54, no. 3, pp. 356–369, Sep. 2019, doi: 10.1111/ejed.12345.
- [33] A. I. Suryani, A. R. Wulan, and I. R. Suwarna, "Validation of science process skills assessment instrument through Rasch model analysis," *J. Phys. Conf. Ser.*, vol. 1806, no. 1, p. 012072, 2021.
- [34] B. D. Wright and J. M. Linacre, "Reasonable mean-square fit values." Accessed: Oct. 23, 2025. [Online]. Available: <https://rasch.org/rmt/rmt83b.htm>
- [35] G. Willis, *Cognitive Interviewing*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc., 2005. doi: 10.4135/9781412983655.
- [36] J. M. Linacre, "What do infit and outfit, mean-square and standardized mean?," *Rasch Meas. Trans.*, vol. 16, no. 2, p. 878, 2002.
- [37] J. M. Linacre, "Investigating rating scale category utility.," *J. Outcome Meas.*, vol. 3, no. 2, pp. 103–22, 1999, [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10204322>
- [38] D. Andrich, I. Marais, and S. Humphry, "Using a Theorem by Andersen and the Dichotomous Rasch Model to Assess the Presence of Random Guessing in Multiple Choice Items," *J. Educ. Behav. Stat.*, vol. 37, no. 3, pp. 417–442, Jun. 2012, doi: 10.3102/1076998611411914.
- [39] M. Wilson, *Constructing Measures*. Routledge, 2004. doi: 10.4324/9781410611697.
- [40] J. Glück, "Measuring Wisdom: Existing Approaches, Continuing Challenges, and New Developments," *Journals Gerontol. Ser. B*, vol. 73, no. 8, pp. 1393–1403, Oct. 2018, doi: 10.1093/geronb/gbx140.
- [41] L. Darling-Hammond, M. Hyler, and M. Gardner, "Effective Teacher Professional Development," Jun. 2017. doi: 10.54300/122.311.
- [42] M. Thurlings and P. den Brok, "Learning outcomes of teacher

- professional development activities: a meta-study,” *Educ. Rev.*, vol. 69, no. 5, pp. 554–576, Oct. 2017, doi: 10.1080/00131911.2017.1281226.
- [43] J. . Stronge, *Qualities of effective teachers*, 3rd ed. ASCD, 2018.
- [44] A. D. Wu and K. Ercikan, “Using Multiple-Variable Matching to Identify Cultural Sources of Differential Item Functioning,” *Int. J. Test.*, vol. 6, no. 3, pp. 287–300, Sep. 2006, doi: 10.1207/s15327574ijt0603_5.
- [45] R. Zwick, “A Review of ETS Differential Item Functioning Assessment Procedures: Flagging Rules, Minimum Sample Size Requirements, and Criterion Refinement,” *ETS Res. Rep. Ser.*, vol. 2012, no. 1, Jun. 2012, doi: 10.1002/j.2333-8504.2012.tb02290.x.
- [46] R. D. Penfield and T. C. M. Lam, “Assessing Differential Item Functioning in Performance Assessment: Review and Recommendations,” *Educ. Meas. Issues Pract.*, vol. 19, no. 3, pp. 5–15, Sep. 2000, doi: 10.1111/j.1745-3992.2000.tb00033.x.
- [47] J. D. Webster, “Self-Report Wisdom Measures,” in *The Cambridge Handbook of Wisdom*, Cambridge University Press, 2019, pp. 297–320. doi: 10.1017/9781108568272.015.
- [48] P. Grossman, C. Compton, D. Igra, M. Ronfeldt, E. Shahan, and P. W. Williamson, “Teaching Practice: A Cross-Professional Perspective,” *Teach. Coll. Rec. Voice Scholarsh. Educ.*, vol. 111, no. 9, pp. 2055–2100, Sep. 2009, doi: 10.1177/016146810911100905.
- [49] J. Lave and E. Wenger, *Situated Learning*. Cambridge University Press, 1991. doi: 10.1017/CBO9780511815355.
- [50] J. Loughran, “Professionally Developing as a Teacher Educator,” *J. Teach. Educ.*, vol. 65, no. 4, pp. 271–283, Sep. 2014, doi: 10.1177/0022487114533386.
- [51] L. E. Kim and K. Asbury, “‘Like a rug had been pulled from under you’: The impact of COVID-19 on teachers in England during the first six weeks of the UK lockdown,” *Br. J. Educ. Psychol.*, vol. 90, no. 4, pp. 1062–1083, Dec. 2020, doi: 10.1111/bjep.12381.