

Adaptation and Validation of The School Bystander Behaviour Scale (SBBS) Instrumen for Junior High School Students: Rasch Model Analysis

[Adaptasi dan Validasi Instrumen School Bystander Behaviour Scale (SBBS) untuk Siswa SMP: Analisis Rasch Model]

Diky Zulkiflismin Zuhri¹⁾, Widayastuti^{*,2)}

¹⁾Program Studi Psikologi, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾Program Studi Psikologi, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi : wiwid@umsida.ac.id

Abstract. *The School Bystander Behavior Scale (SBBS) is one of the instruments that can measure the tendency of bystander students involved in bullying. This study aims to adapt the School Bystander Behavior Scale (SBBS) instrument into Indonesian and analyze the quality of its psychometric properties. Rasch model is used as the testing method on data obtained from 251 respondents who are 7th, 8th, and 9th grade students on junior high school which was selected using proportionate stratified random sampling. The results of the study indicate that the adapted instrument has sufficient unidimensionality and good reliability, but some items have low validity. Therefore, the instrument requires further review to improve its psychometric properties and performance.*

Keywords – instrument adaptation; rasch model; School Bystander Behaviour Scale (SBBS)

Abstrak. *Instrumen School Bystander Behaviour Scale (SBBS) merupakan salah satu instrumen yang dapat mengukur kecenderungan sikap siswa bystander yang terlibat dalam bullying. Penelitian ini bertujuan untuk melakukan adaptasi dari instrumen School Bystander Behaviour Scale (SBBS) ke dalam Bahasa Indonesia dan menganalisis kualitas sifat psikometriknya. Metode pengujian menggunakan rasch model terhadap data yang diperoleh dari 251 responden yang merupakan siswa SMP kelas 7, 8, dan 9 yang dipilih dengan teknik proportionate stratified random sampling. Hasil penelitian menunjukkan bahwa instrumen hasil adaptasi memiliki unidimensionalitas yang cukup dan reliabilitas yang baik, namun beberapa itemnya memiliki validitas yang rendah. Oleh karena itu, instrumen memerlukan pengkajian lebih lanjut agar dapat memiliki sifat psikometrik dan memberikan performa yang lebih baik.*

Kata Kunci – adaptasi instrumen; rasch model; School Bystander Behaviour Scale (SBBS)

I. PENDAHULUAN

Bystander effect merujuk pada fenomena menurunnya keinginan individu untuk membantu orang lain ketika terdapat *bystander* lain yang juga bersikap pasif [1]. Kita sering mendapatkan contoh dari fenomena ini, baik yang terjadi depan mata kita sendiri atau yang kita saksikan melalui media sosial dan elektronik. Terdapat empat faktor utama yang melatarbelakangi terjadinya *bystander effect* yakni: (1) kesadaran diri; (2) isyarat sosial; (3) mekanisme *blocking*; (4) difusi tanggung jawab [2]. Fenomena *bystander effect* dapat terjadi pada situasi darurat maupun nondarurat. Respon *bystander* pada situasi tersebut dapat memberikan dampak yang bervariasi, namun dampak yang paling ekstrim adalah hilangnya nyawa, seperti pada kasus pemerkosaan dan pembunuhan Kitty Genovese pada tahun 1964 [1]. Penelitian menunjukkan bahwa *bystander effect* juga terjadi dalam *bullying* pada siswa sekolah menengah pertama (SMP) [3]. Remaja pada usia ini sedang dalam proses pencarian jati diri, sehingga ketika terdapat *bullying*, tak jarang mereka bingung bagaimana harus bersikap. Selain itu, *bystander* yang berasal dari pihak teman sebaya seringkali memiliki anggapan bahwa *bullying* adalah hal yang wajar dan tidak akan berdampak negatif [4]. Anggapan mengenai tidak adanya dampak negatif ini merupakan sesuatu yang keliru. Hal ini didukung oleh studi bahwa seorang remaja yang menjadi korban *bullying* dapat menimbulkan dampak negatif berupa perasaan tidak percaya diri, terisolasi, depresi, atau bahkan bunuh diri [5]. Hal ini diperkuat dengan penelitian yang menunjukkan bahwa meningkatnya kasus *bullying* selama periode 2020 hingga 2023 berpengaruh positif terhadap tren bunuh diri remaja korban *bullying* [6]. Sebagai generasi penerus, tidak seharusnya siswa SMP bersikap pasif ketika menyaksikan adanya *bullying* yang menimpakan teman sekelasnya.

Studi terdahulu menunjukkan bahwa siswa SMP, yang notabenenya merupakan remaja dengan rentang usia 12-15 tahun, rentan menjadi korban atau pelaku *bullying* [7]. Hal ini didukung dengan penelitian yang menunjukkan bahwa 188 siswa SMP dari 3 sekolah di Surabaya, Jawa Timur, Indonesia pernah menjadi pelaku, korban, ataupun keduanya [8]. Penelitian lebih lanjut menunjukkan bahwa intensitas *bullying* turut dipengaruhi oleh persepsi negatif pelaku terhadap *bystander* yang berpikir bahwa *bystander* akan membiarkan pelaku atau justru mendukung perilaku *bullying* [3].

Adanya keterkaitan basis sensitivitas moral, lepasnya moralitas, dan *self-efficacy* menentukan apakah siswa akan bertindak sebagai *bystander*, *defender*, atau *probully* pada siswa SMP. Penelitian menunjukkan bahwa basis sensitivitas moral berkontribusi negatif terhadap lepasnya moralitas, sehingga siswa yang memiliki basis sensitivitas moral yang rendah akan mengakibatkan tindakan diluar moral yang justru akan cenderung mendukung pelaku *bullying*. Jika basis sensitivitas moral tinggi namun memiliki *self-efficacy* yang rendah, siswa akan cenderung bertindak pasif sebagai *bystander*. Sebaliknya, jika sensitivitas moral tinggi dan disertai dengan *self-efficacy* yang juga tinggi, siswa akan cenderung bertindak sebagai *defender* [9]. Penelitian juga menunjukkan bahwa individu dan kelas yang mendukung norma anti-*bullying* akan cenderung membela korban *bullying*. Selain itu, diungkapkan bahwa perempuan cenderung lebih mendukung norma anti-*bullying*, serta memiliki kelepasan moralitas lebih rendah daripada laki-laki [10].

Penelitian terdahulu mengungkapkan realita bahwa adanya respon aktif *bystander* berkontribusi terhadap intensitas *bullying* pada siswa SMP. Perilaku *bullying* semakin berkurang jika semakin banyak teman sebaya yang membela korban [11]. Hal ini menunjukkan bahwa respon pasif *bystander* perlu diubah agar dapat mendukung perilaku anti-*bullying* [12]. Selain itu, penelitian menunjukkan bahwa korban *bullying* seringkali mencari bantuan pihak lain [13]. Dalam hal ini, bantuan atau kontribusi positif *bystander* seharusnya diberikan kepada korban *bullying*. Hal ini dikarenakan penelitian menunjukkan bahwa korban *bullying* yang mendapat bantuan dari *bystander*, yang berasal dari pihak teman sebaya, memberikan korban persepsi positif mengenai lingkungan sekolah sehingga akan turut meningkatkan *subjective well-being* korban [14]. Penelitian juga menunjukkan bahwa adanya dukungan sosial *bystander* dapat meningkatkan kepercayaan diri korban *bullying* [15].

Sebagai upaya mengukur fenomena *bystander effect* siswa dalam *bullying*, para peneliti telah merancang instrumen-instrumen penelitian. Instrumen *Bystander Intervention Model For Bullying And Sexual Harassment* dirancang dengan tujuan untuk mengukur proses struktural intervensi *bystander* [16]. Instrumen *Bullying Participant Behavior Questionnaire* (BPBQ) disusun untuk menentukan kecenderungan sikap siswa dengan klasifikasi sebagai *bully*, *assistant*, *victim*, *defender*, dan *outsider* di US [17]. *Participant Role Questionnaire* (PRQ) juga disusun untuk menentukan kecenderungan sikap siswa, akan tetapi memiliki susunan yang berbeda dengan BPBQ, yakni *bully*, *assistant*, *reinforcer*, *defender*, dan *outsider* [18]. Instrumen yang diciptakan oleh para peneliti sebelumnya menunjukkan kecenderungan sikap dari perspektif siswa secara keseluruhan dalam kasus *bullying*. Meski demikian, hampir tidak ada instrumen yang meninjau kecenderungan sikap dalam kasus *bullying* yang secara spesifik berdasarkan perspektif para siswa *bystander*.

Instrumen *School Bystander Behaviour Scale* (SBBS) adalah instrumen yang dapat mengukur kecenderungan sikap siswa dalam kasus *bullying* yang secara spesifik berdasarkan perspektif para siswa *bystander*. Instrumen ini dapat mengukur dan mengklasifikasikan kecenderungan sikap siswa sebagai *bystander* pasif, *defender*, dan *probully*. Klasifikasi pada instrumen ini didasarkan dari temuan sebelumnya yang menunjukkan bahwa hubungan antara basis sensitivitas moral, lepasnya moralitas, dan *self-efficacy* menentukan bagaimana kecenderungan sikap siswa *bystander*. Instrumen ini dikembangkan di Swedia dengan 347 remaja (141 laki-laki dan 206 perempuan) dengan rentang usia 15-20 tahun sebagai partisipan. Partisipan yang dipilih adalah mereka yang pernah menyaksikan atau terlibat dalam *bullying* di sekolah. Instrumen ini memiliki 8 aitem soal yang dengan menggunakan 5-basis point skala likert sebagai pilihan jawaban. *Exploratory Factor Analysis* menunjukkan bahwa aitem-aitem dalam instrumen ini memiliki *loading factor* yang dapat diterima yakni, *probully* (0,53-0,87), *bystander* (0,55-0,90), dan *defender* (0,70-0,80) [9].

Adaptasi instrumen *School Bystander Behaviour Scale* (SBBS) pernah dilakukan di Asturias, Spanyol dengan 1,018 siswa sebagai partisipan yang memiliki rentang usia 9-11 tahun. Adaptasi dilakukan dengan mengikutsertakan siswa yang belum pernah menyaksikan *bullying*, menambahkan 2 aitem (1 aitem *bystander* dan 1 aitem *defender*) sehingga jumlah keseluruhan aitem menjadi 10, dan mengubah format respon menjadi dikotomi (Ya/Tidak). *Confirmatory Factor Analysis* menunjukkan *loading factor* yang dapat diterima, yakni *probully* (0,70-0,89), *bystander* (0,63-0,83), dan *defender* (0,62-0,71). Reliabilitas final yang dimiliki *probully* = 0,85, *bystander* = 0,81, dan *defender* = 0,75 [19].

Instrumen *School Bystander Behaviour Scale* (SBBS) belum pernah menjadi subjek penelitian di Indonesia. Hal ini menjadi penting karena bahasa dapat mempengaruhi interpretasi, pengambilan keputusan, dan strategi [20]. Oleh karena itu, skala ini memerlukan adaptasi agar sesuai dengan konteks bahasa Indonesia. Selain itu, penelitian adaptasi instrumen SBBS belum pernah menggunakan pendekatan analisis Rasch Model. Penelitian dengan pendekatan rasch model penting dilakukan karena rasch model dapat mengungkap indeks reliabilitas item setiap level,

reliabilitas responden, unidimensionalitas, dan bias pada item [21]. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan instrumen dengan metode adaptasi untuk mengukur kecenderungan sikap siswa *bystander* pada jenjang pendidikan SMP di Indonesia dengan metode analisis rasch model.

II. METODE

Penelitian ini merupakan penelitian adaptasi alat ukur dengan pendekatan kuantitatif menggunakan desain forward translation. Untuk melakukan adaptasi alat ukur terdapat beberapa tahap yang harus dipenuhi, yakni: (1) translasi; (2) sintesis; (3) *expert committee review* [22]. Pendekatan kuantitatif adalah pendekatan yang menggunakan angka-angka dan analisis statistik [23]. Adapun variabel yang terdapat pada penelitian ini adalah *bystander effect*.

Populasi pada penelitian ini adalah siswa SMP Negeri 2 Pungging yang memiliki jumlah sebanyak 741 siswa. Batasan jumlah sampel pada penelitian ini ditentukan berdasarkan ketentuan 5% *margin of error* yang dirumuskan. Adapun populasi memiliki tingkatan strata kelas 7, 8, dan 9, sehingga teknik pengambilan sampel yang digunakan adalah *proportionate stratified random sampling*. Teknik ini sesuai karena target populasi bersifat heterogen dan memiliki strata yang proporsional [24]. Sampel dari masing-masing strata kelas dipilih secara acak dengan pembagian sebagai berikut :

Kelas	Jumlah Siswa	Jumlah Sampel
7	251	87
8	244	80
9	246	84
Total	741	251

Tabel 1. Pembagian jumlah sampel

Ekstraksi data dilakukan dengan instrumen yang disusun oleh peneliti dengan metode adaptasi dari instrumen pengukuran yang disusun oleh Suárez-García, dkk. berdasarkan instrumen asli yang dirancang Thornberg & Jungert yang bernama *School Bystander Behaviour Scale* (SBBS). Instrumen ini dapat mengkategorisasikan tiga jenis perilaku *bystander* dalam *bullying* yakni *bystander* pasif, *defender*, dan *probully*. Instrumen SBBS memenuhi standar dengan memiliki *model-fit* ($N = 347$; $CFI = .92$; $S-BSS = 258.44$ $df = 144$; $RMSEA = .048$, 90% CI [0.04, 0.06]) dengan *loading factor* *bystander* pasif (0,63-0,83), *defender* (0,62-0,71), dan *probully* (0,70-0,89) *probully* [19]. Adaptasi instrumen pada penelitian ini menggunakan butir respon rating scale dengan 4 poin skala Likert dengan skor 1 sebagai nilai minimum dan skor 4 sebagai nilai maksimum. Butir respon diisi dengan ceklist terhadap 4 pilihan pernyataan Sangat Sesuai (SS), Sesuai (S), Tidak Sesuai (TS), dan Sangat Tidak Sesuai (STS).

Aspek dari instrumen yang diadaptasi terbagi menjadi 3, yakni *bystander* pasif, *defender*, dan *probully*. Ketiga aspek ini terdiri dari beberapa aitem sebagai indikator perilaku yang ditunjukkan oleh tabel di bawah ini :

Aspek	Indikator perilaku
<i>Bystander</i> pasif	1, 7, 8
<i>Defender</i>	2, 3, 4, 6
<i>Probully</i>	5, 9, 10

Tabel 2. Aspek dan indikator perilaku pada aitem

Adapun instrumen final hasil adaptasi adalah sebagai berikut :

Angket "Skala Perilaku Siswa Bystander" Untuk Siswa Sekolah Menengah Pertama

Perundungan (*Bullying*) adalah istilah untuk menggambarkan situasi dimana seorang siswa menjadi korban berbagai jenis agresi dari waktu ke waktu di sekolah, yang dilakukan oleh siswa lain. Jenis agresi tersebut seperti mengolok-olok, menghina, mengancam, mengucilkan, menyembunyikan barang-barangnya, dst. Dalam situasi perundungan, korban berada dalam posisi yang lebih lemah dibandingkan pelaku (misalnya, lebih pemalu atau memiliki lebih sedikit teman).

*Di tahun ajaran ini, pernahkah kamu melihat teman sekelasmu menjadi korban perundungan?
(lingkari pilihanmu)*

Jika kamu pernah menemui temanmu menjadi korban perundungan di kelas pada tahun ajaran ini, “apa yang kamu lakukan saat hal itu terjadi?”, tandai pilihan dibawah. Jika kamu belum pernah melihat temanmu menjadi korban

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.

Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

perundungan di kelas pada tahun ajaran ini, "apa yang akan kamu lakukan seandainya kamu melihat temanmu menjadi korban perundungan?", tolong tandai pilihan dibawah.

Setelah membaca penjelasan di atas, apakah kamu bersedia untuk berpartisipasi dalam angket ini? (lingkari pilihanmu) Ya / Tidak

Berilah tanda centang (✓) pada kolom di samping yang sesuai dengan dirimu

Keterangan : SS (Sangat Sesuai), S (Sesuai), TS (Tidak Sesuai), STS (Sangat Tidak Sesuai)

No	Pernyataan	SS	S	TS	STS
1	Mengambil peran untuk membela teman sekelas saya dan membuat pelaku perundungan berhenti melakukannya.				
2	Bicara pada pelaku perundungan terlebih dahulu agar mereka berhenti.				
3	Memberitahu guru.				
4	Mendorong/memotivasi korban untuk bercerita kepada orang dewasa (anggota keluarga, guru) tentang apa yang terjadi.				
5	Saya ikut serta dan mulai mem-bully siswa tersebut.				
6	Bicara pada teman sekelas lainnya untuk mencoba menghentikan apa yang terjadi.				
7	Tidak ada. Saya tetap melakukan hal yang sedang saya kerjakan karena menurut saya, apa yang terjadi tidak ada hubungannya dengan saya.				
8	Saya menjauh.				
9	Menontonnya, karena itu lucu dan menghibur.				
10	Tertawa dan mendukung para penindas.				

Tabel 3. Instrumen final hasil adaptasi

Teknik analisis data pada penelitian ini menggunakan rasch model untuk mendapatkan informasi mengenai kualitas instrumen yang digunakan. Peneliti menggunakan rasch model dikarenakan rasch model memenuhi lima standar kriteria, yakni; (1) ukuran linear dengan interval yang sama; (2) proses estimasi yang tepat; (3) dapat menemukan aitem yang tidak tepat (*misfits*) atau tidak umum (*outliers*); (4) dapat mengatasi data yang hilang; (5) menghasilkan pengukuran yang *replicable* [25]. Adapun rekapitulasi data mentah dilakukan dengan bantuan *software* Microsoft Excel dan analisis data menggunakan rasch model dilakukan dengan bantuan *software* Winstep

III. HASIL PENELITIAN

1. Hasil Penelitian

A. Validitas

Uji validitas instrumen dilaksanakan dengan menganalisis hasil uji unidimensionalitas, *item misfit order*, *Wright map* (guna menunjukkan distribusi *item-person*), serta *category function*.

a. Unidimensionalitas

Hasil uji unidimensionalitas dalam rasch model menunjukkan apakah instrumen dapat mengukur konstruk yang seharusnya diukur. Dalam hal ini, instrumen yang memiliki unidimensionalitas yang baik merupakan instrumen yang dapat mengukur hanya satu konstruk yang dimaksud saja [26].

INPUT: 251 PERSON 10 ITEM REPORTED: 251 PERSON 10 ITEM 4 CATS WINSTEPS 5.1.7.0

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units			
	Eigenvalue	Observed	Expected
Total raw variance in observations	= 24.9414	100.0%	100.0%
Raw variance explained by measures	= 14.9414	59.9%	59.5%
Raw variance explained by persons	= 2.2944	9.2%	9.1%
Raw Variance explained by items	= 12.6470	50.7%	50.3%
Raw unexplained variance (total)	= 10.0000	40.1% 100.0%	40.5%
Unexpln variance in 1st contrast	= 2.9266	11.7% 29.3%	
Unexpln variance in 2nd contrast	= 1.1923	4.8% 11.9%	
Unexpln variance in 3rd contrast	= 1.1039	4.4% 11.0%	
Unexpln variance in 4th contrast	= 1.0130	4.1% 10.1%	
Unexpln variance in 5th contrast	= .8636	3.5% 8.6%	

Tabel 4. Hasil pengujian unidimensionalitas

Hasil uji unidimensionalitas pada tabel 4 menunjukkan *raw variance explained by measures* memiliki nilai = 59,9% (50-60%). *Raw variance explained by measure* merupakan proporsi dari varians total dalam data yang dapat

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.

Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

dijelaskan oleh pengukuran yang telah dilakukan [27]. Nilai persentase sebesar 59,9% menandakan bahwa instrumen cukup dapat mengukur konstruk yang seharusnya diukur. Nilai ini juga menunjukkan bahwa varians total data mengenai konstruk yang seharusnya diukur cukup banyak dalam pengukuran yang telah dilakukan. Selain itu, terdapat *unexplained variance in 1st-5th contrast* yang menunjukkan residual berupa varians lain yang turut terukur dengan nilai yang tidak boleh melebihi 15% [25]. Nilai *unexplained variance pada 1st contrast* = 11,7%, *2nd contrast* = 4,8%, *3rd contrast* = 4,4%, *4th contrast* = 4,1%, dan *5th contrast* = 3,5%. Berdasarkan hasil tersebut, tidak terdapat nilai yang melebihi 15% pada *unexplained variance in 1st-5th contrast*. Ini berarti varians lain yang turut diukur oleh instrumen masih dalam batas yang dapat diterima. Lebih lanjut, hal ini juga menunjukkan bahwa instrumen memiliki *local independence* yang masih cukup tidak terpengaruh dengan varians lain di luar konstruk.

b. Item Misfit Order

Hasil pada *item misfit order* yang bertujuan untuk mengevaluasi kecocokan (*fit*) dari setiap aitem dalam instrumen ini [28]. *Outfit MNSQ* pada *item misfit order* menunjukkan deviasi *outlier* aitem dengan nilai $range 0,5 \leq x \leq 1,5$ pada aitem individual [29].

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT MNSQ	PTMEASUR-AL ZSTD	ZSTD CORR.	EXACT EXP.	MATCH OBS%	ITEM EXP%
4	785	251	-1.58	.09	1.38	4.04	1.47	4.76	A .15	.38	46.2
3	812	251	-1.81	.09	1.24	2.61	1.28	2.91	B .22	.37	53.8
1	409	251	1.21	.10	1.10	1.19	1.17	1.79	C .17	.42	57.8
5	344	251	1.94	.12	1.16	1.43	.94	-.50	D .52	.38	68.9
9	401	251	1.29	.10	.99	-.04	.87	-1.43	E .68	.41	66.9
10	364	251	1.68	.11	.98	-.22	.88	-1.16	e .54	.39	66.5
2	804	251	-1.74	.09	.81	-2.28	.93	-.82	d .14	.37	62.5
6	779	251	-1.53	.09	.80	-2.50	.85	-1.79	c .30	.39	62.9
7	552	251	.10	.08	.81	-2.57	.82	-2.38	b .60	.44	55.0
8	504	251	.44	.09	.76	-3.39	.77	-3.17	a .60	.44	57.0
MEAN	575.4	251.0	.00	.10	1.00	-.17	1.00	-.18		59.8	53.8
P.SD	188.6	.0	1.45	.01	.20	2.36	.22	2.39		6.7	6.5

Tabel 5. Hasil pengujian *item misfit order*

Berdasarkan hasil analisis *item misfit order* pada tabel 5, menunjukkan bahwa nilai *Outfit MNSQ* pada setiap aitem berada dalam rentang $0,5 \leq x \leq 1,5$. *Outfit MNSQ* menunjukkan besar statistik *outlier* (respon yang ekstrem dan tidak terduga sehingga jauh dari ekspektasi) yang bersifat deskriptif pada rasch model [25]. Dikarenakan semua nilai *Outfit MNSQ* berada dalam rentang standar, hal ini memiliki arti bahwa semua aitem pada instrumen SBBS yang telah diadaptasi memiliki *outlier* yang bersifat deskriptif yang dapat diterima oleh rasch model. Adapun aitem yang memiliki nilai *Outfit MNSQ* terbesar terdapat pada aitem D3 dengan nilai = 1,47.

Outfit ZSTD bertujuan untuk mengidentifikasi *outlier* aitem yang tidak *fit* dengan model rasch yang mana memiliki standar nilai $-2.0 \leq x \leq 2.0$ [30]. *Outfit ZSTD* bersifat inferensial sehingga *Outfit ZSTD* memberikan pemahaman mengenai signifikansi terhadap aitem yang tidak *fit* dengan rasch model [25]. Dalam hal ini, hasil penelitian menunjukkan bahwa terdapat 4 aitem yang tidak *fit* dengan model rasch, yakni aitem B2 (-2,38), B3 (-3,17), D2 (2,91) dan D3 (4,76). Oleh karena itu, aitem B2, B3, D2, dan D4 memerlukan kajian ulang agar dapat memenuhi kriteria *Outfit ZSTD*.

Nilai yang ditunjukkan *point measure correlation* menunjukkan daya beda aitem dengan *range* nilai standar $= 0,4 \leq x \leq 0,85$ [31]. Hasil *point measure correlation* menunjukkan bahwa terdapat beberapa aitem pada instrumen penelitian yang tidak memenuhi standar, yakni aitem B1 (0,17), D1 (0,14), D2 (0,22), D3 (0,15), dan D4 (0,30). Hal ini menandakan bahwa aitem-aitem tersebut memiliki daya beda yang kurang baik sehingga kurang dapat membedakan tingkat kemampuan responden.

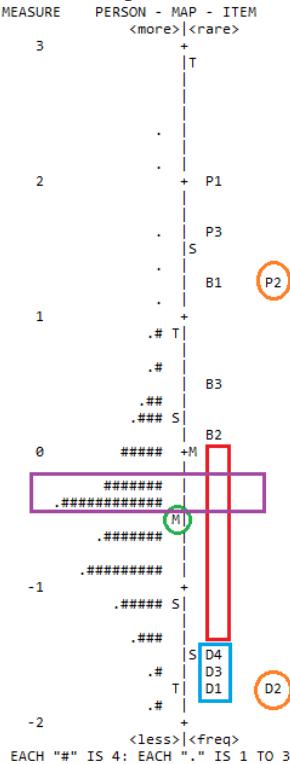
JMLE Measure memberikan petunjuk untuk menentukan mana aitem yang paling mudah dan paling sulit disetujui oleh responden. Aitem dengan nilai logit yang tinggi memiliki arti bahwa aitem tersebut lebih sulit untuk disetujui. Hal ini berkorelasional dengan kemampuan responden di mana responden yang memiliki kemampuan lebih tinggi, cenderung dapat memberikan respon yang memiliki skor tinggi pada aitem yang sulit disetujui [25]. Berdasarkan data *JMLE Measure* yang didapatkan menunjukkan bahwa nilai rata-rata standar deviasi logit aitem = +145. Adapun nilai logit tertinggi terdapat pada aitem P1 = +194 dan nilai logit terendah terdapat pada aitem D2 = -181. Hal ini memiliki arti bahwa aitem P1 adalah aitem paling sulit disetujui, sedangkan aitem D2 merupakan aitem yang paling mudah disetujui.

c. Wright Map

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.

Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

Hasil *Wright map* memuat distribusi *item-person* yang merupakan visualisasi pemetaan untuk mengetahui distribusi tingkat kesulitan aitem dan kemampuan individu yang mana hal ini juga menunjukkan seberapa baik hubungan di antara keduanya. Bagian kanan *map* memuat aitem yang diujikan, sedangkan bagian kiri memuat responden. Sumbu vertikal menunjukkan tingkat kesulitan aitem yang sesuai dengan tingkat kemampuan responden pada sumbu horizontal. Aitem-aitem yang sulit dan membutuhkan kemampuan tinggi responden terletak di bagian atas *map*, sedangkan aitem-aitem yang mudah dan tidak membutuhkan kemampuan tinggi responden terletak di bagian bawah *map*. Selaras dengan tingkat kesulitan aitem, responden yang memiliki kemampuan tinggi terletak di bagian atas *map*, sedangkan responden yang memiliki kemampuan rendah terletak di bagian bawah *map* [32].



Gambar 1. Wright map

Gambar 1 menunjukkan bahwa terdapat sebaran aitem yang tidak proporsional. Pada kotak merah menunjukkan bahwa terdapat *gap* berupa kekosongan aitem akan tetapi terdapat responden yang berada dalam *range* kekosongan tersebut. Hal ini mengakibatkan hasil pengukuran kemampuan beberapa responden menjadi kurang terwakilan karena aitem yang tersedia belum mencukupi. Sebagai contoh yakni dalam kotak ungu, terdapat beberapa responden yang terletak pada *range* kekosongan tersebut. Kemampuan responden yang telah diukur dalam kotak ungu menjadi kurang detail dan akurat akibat tidak adanya aitem yang mewakili kemampuan responden tersebut. Dalam mengatasi hal ini, diperlukan adanya aitem untuk mengisi *gap* yang ada. Selain itu, terdapat aitem yang memiliki tingkat kesulitan yang tidak jauh berbeda, yakni aitem D1, D3, dan D4. Selain itu, terdapat pula aitem D2 yang memiliki tingkat kesulitan yang hampir sama dengan aitem D1 dan D3, dan aitem P2 yang hampir sama dengan B1.

Hasil *Wright map* pada gambar 1 juga menunjukkan sebaran *item-person* yang tidak merata dengan mayoritas responden terkumpul di bagian bawah *map*. Hal ini menunjukkan bahwa sedikit responden yang dapat menjawab dengan benar. Adapun responden rata-rata hanya dapat menjawab 2 soal benar (skor 4) dari 10 soal yang diujikan. Jika diperhatikan dengan lebih detail, terdapat 1 responden yang memiliki kemampuan sangat tinggi, yakni responden ke-232 yang terletak pada bagian paling atas *map* dengan nilai logit +243. Hal ini menandakan bahwa responden ke-232 dapat menjawab mayoritas soal dengan benar. Terdapat pula 5 responden yang memiliki kemampuan sangat rendah, yakni responden ke-107, ke-176, ke-195, ke-220, dan ke-226 yang terletak pada bagian paling bawah *map* dengan nilai logit = -187. Adapun responden ke-226 hanya dapat menjawab benar pada 2 soal, responden ke-176 dan ke-195 hanya dapat menjawab benar pada 1 soal, dan responden ke-107 dan ke-220 tidak dapat menjawab benar pada semua soal.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL MNSQ	EXACT CORR.	MATCH EXP.	EXACT OBS%	MATCH EXP%
MEAN	22.9	10.0	-.45	.48	.99	-.25	1.00	-.22		59.8	53.8
P.SD	2.9	.0	.66	.02	.84	1.56	.90	1.48		19.3	5.2

Tabel 6. Hasil pengujian *person measure*

Dari tabel 6 diketahui bahwa nilai logit rata-rata responden = -45 yang mana hal ini lebih rendah dari nilai logit rata-rata aitem (0,0). Nilai ini juga sesuai dengan gambar 1 di mana *nilai mean person* (*M*) berada di bawah *mean* aitem (+*M*). Hal ini menandakan bahwa rata-rata responden memiliki kemampuan yang lebih rendah dibandingkan tingkat kesulitan aitem.

d. Category Function

Instrumen yang digunakan dalam penelitian ini menggunakan *rating scale*, sehingga perlu untuk menilai kualitas butir respon melalui pengujian *category function*. *Category function* menilai kualitas butir respon rating scale yang berkorespondensi dengan tingkat kemampuan responden [33].

SUMMARY OF CATEGORY STRUCTURE. Model="R"										
CATEGORY LABEL	OBSERVED			OBSVSD SAMPLE		INFIT OUTFIT		ANDRICH THRESHOLD	CATEGORY MEASURE	
	SCORE	COUNT	%	AVRG E	EXPECT	MNSQ	MNSQ			
1 1	730	29	-1.99	-1.94	1.01	1.08		NONE	(-2.71)	1
2 2	701	28	-.97	-.92	.82	.82		-1.45	-.88	2
3 3	694	28	.82	.54	.64	.68		-.16	.82	3
4 4	385	15	1.12	1.40	1.46	1.47		1.60	(2.82)	4

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

Tabel 7. Hasil pengujian *category function*

Kolom *category measure* pada tabel 7 menunjukkan bahwa butir respon memiliki nilai *category measure* yang semakin naik dengan butir respon skor 1 memiliki nilai = -2,71, skor 2 = -0,88, skor 3 = 0,82, dan skor 4 = 2,82. Nilai ini menunjukkan bahwa kesesuaian respon yang diberikan responden meningkat sesuai dengan rata-rata tingkat kemampuan yang dimiliki, sehingga semakin tinggi kemampuan yang dimiliki responden, maka responden akan semakin cenderung menjawab butir dengan skor 4. Selain itu, nilai dari kolom *category measure* menunjukkan bahwa butir respon skor 1 hingga skor 4 tertata secara runut.

Kolom *observed score count %* di tabel 7 menunjukkan bahwa semua butir respon pada instrumen yang diujikan telah terisi. Butir respon dengan skor 1 paling banyak dipilih dengan persentase sebesar 29% dari jumlah keseluruhan, sedangkan butir respon dengan skor 4 paling sedikit dipilih dengan persentase 15%. Ini menunjukkan bahwa responden yang memiliki tingkat kemampuan yang tinggi lebih sedikit daripada responden yang memiliki tingkat kemampuan yang rendah.

Observed Average Measures for PERSON (unscored) (by Observed Category)						
-5	-3	-1	1	3	5	
						NUM ITEM
		1 2 34				5 P1
		1 2 3 4				10 P3
		1 2 34				9 P2
		1324				1 B1
		1 2 3 4				8 B3
		1 2 3 4				7 B2
		1 23 4				6 D4
		2134				4 D3
		1 234				2 D1
		1 234				3 D2
		1				NUM ITEM
-5	-3	-1	1	3	5	
						PERCENTILE
		12 334 221 1				
		55 51 709 805 055 121 1 1				PERSON
		T S M S T				
		0 10 40 70 80 90 99				

Gambar 2. Hasil pengujian *empirical item-category function*

Hasil uji *empirical item-category function* pada gambar 2 menunjukkan secara detail bahwa terdapat beberapa pilihan jawaban dengan susunan skor yang tidak berurutan. Aitem yang memiliki susunan skor yang tidak berurutan adalah aitem B1 (1-3-2-4) dan aitem D3 (2-1-3-4). Hal ini menunjukkan arti bahwa aitem B1 dan aitem D3 merupakan aitem yang memiliki respon yang membingungkan responden. Selain itu, mayoritas aitem memiliki *range* skor yang tidak cukup jauh sehingga kurang dapat membedakan tingkat kemampuan responden.

e. Person Misfit Order

Hasil pengujian *person misfit order* menunjukkan responden yang tidak *fit* dengan rasch model. Untuk menentukan *fit* atau tidaknya responden, nilai *Infit* MNSQ dan *Outfit* MNSQ digunakan sebagai acuan. Baik *Infit* MNSQ dan *Outfit* MNSQ harus berada dalam rentang standar ($-1 \leq x \leq 1$) agar seorang responden dinyatakan *fit* dengan rasch model [34].

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	EXACT EXP.	MATCH OBS%	EXACT EXP%
147	26	10	.22	.46 6.43	6.38 7.28	6.97 A-.69	.75	20.0	46.6			
207	26	10	.22	.46 6.32	6.31 6.98	6.77 B-.51	.75	0	46.6			
246	21	10	-.87	.48 4.05	4.11 4.04	3.91 C .20	.75	50.0	58.4			
232	35	10	2.43	.58 3.36	2.96 3.97	2.54 D .03	.58	40.0	70.0			
228	19	10	-1.35	.50 2.91	2.84 3.77	3.28 E .29	.73	60.0	59.8			
129	25	10	.01	.46 3.59	3.99 3.75	4.15 F .09	.75	30.0	45.3			
1	27	10	.43	.46 3.06	3.39 3.71	4.08 G-.06	.74	40.0	51.7			
164	34	10	2.12	.55 2.00	1.68 3.42	2.50 H-.08	.62	30.0	66.6			
150	24	10	-.21	.46 3.06	3.38 3.39	3.72 I .03	.76	50.0	49.0			
108	25	10	.01	.46 3.11	3.48 3.04	3.39 J .01	.75	20.0	45.3			
22	20	10	-1.10	.49 3.04	3.05 2.45	2.25 K .44	.74	40.0	59.6			
165	26	10	.22	.46 2.83	3.15 2.88	3.21 L .28	.75	10.0	46.6			
53	27	10	.43	.46 2.59	2.82 2.53	2.74 M .23	.74	10.0	51.7			
177	24	10	-.21	.46 2.40	2.58 2.34	2.47 N .44	.76	30.0	49.0			
216	19	10	-1.35	.50 1.93	1.70 2.39	2.05 O .23	.73	20.0	59.8			
209	20	10	-1.10	.49 2.32	2.25 2.03	1.76 P .52	.74	80.0	59.6			
217	20	10	-1.10	.49 2.32	2.25 2.03	1.76 Q .52	.74	80.0	59.6			
86	23	10	-.42	.47 2.31	2.42 2.28	2.34 R .64	.76	40.0	51.0			
141	31	10	1.32	.49 2.27	2.20 1.69	1.31 S .73	.69	40.0	60.4			
244	28	10	.64	.47 1.93	1.87 2.25	2.32 T .15	.73	40.0	54.7			
88	28	10	.64	.47 2.10	2.13 2.24	2.30 U .10	.73	20.0	54.7			
116	27	10	.43	.46 2.21	2.31 2.01	2.03 V .52	.74	20.0	51.7			
16	32	10	1.57	.50 2.17	2.00 2.14	1.77 W .27	.67	40.0	61.7			
189	22	10	-.64	.47 2.17	2.17 1.98	1.85 X .41	.75	20.0	54.3			
3	29	10	.86	.47 1.98	1.91 2.14	2.09 Y .61	.72	50.0	56.9			
11	27	10	.43	.46 1.96	1.94 2.01	2.02 Z .53	.74	50.0	51.7			
14	21	10	-.87	.48 1.95	1.81 1.94	1.72 Z .30	.75	20.0	58.4			
117	26	10	.22	.46 1.81	1.73 1.94	1.93 .58	.75	60.0	46.6			
120	26	10	.22	.46 1.81	1.73 1.94	1.93 .58	.75	60.0	46.6			
127	26	10	.22	.46 1.81	1.73 1.94	1.93 .58	.75	60.0	46.6			
98	23	10	-.42	.47 1.92	1.86 1.90	1.79 .18	.76	30.0	51.0			
132	19	10	-1.35	.50 1.70	1.38 1.91	1.52 .47	.73	40.0	59.8			
226	17	10	-1.87	.52 1.91	1.61 1.23	.55 .71	.69	70.0	64.0			
220	17	10	-1.87	.52 1.68	1.29 1.89	1.31 .42	.69	60.0	64.0			
97	19	10	-1.35	.50 1.88	1.63 1.41	.85 .71	.73	60.0	59.8			
146	30	10	1.09	.48 1.88	1.72 1.45	1.00 .77	.71	50.0	58.2			
156	24	10	-.21	.46 1.70	1.52 1.78	1.65 .61	.76	30.0	49.0			
56	27	10	.43	.46 1.59	1.35 1.75	1.61 .45	.74	40.0	51.7			
99	19	10	-1.35	.50 1.74	1.42 1.63	1.16 .50	.73	40.0	59.8			
237	22	10	-.64	.47 1.66	1.41 1.42	.98 .68	.75	60.0	54.3			
95	20	10	-1.10	.49 1.64	1.32 1.37	.83 .63	.74	70.0	59.6			
6	28	10	.64	.47 1.63	1.39 1.48	1.12 .54	.73	30.0	54.7			
148	24	10	-.21	.46 1.62	1.39 1.63	1.40 .30	.76	20.0	49.0			
82	24	10	-.21	.46 1.55	1.26 1.58	1.32 .34	.76	40.0	49.0			
123	25	10	.01	.46 1.48	1.15 1.58	1.33 .22	.75	40.0	45.3			
186	20	10	-1.10	.49 1.43	.98 1.57	1.14 .58	.74	50.0	59.6			
24	23	10	-.42	.47 1.56	1.26 1.37	.90 .65	.76	40.0	51.0			
153	25	10	.01	.46 1.56	1.29 1.47	1.14 .65	.75	50.0	45.3			
162	18	10	-1.60	.51 1.48	1.02 1.22	.54 .69	.71	60.0	61.2			
221	22	10	-.64	.47 1.46	1.06 1.32	.80 .62	.75	60.0	54.3			
47	23	10	-.42	.47 1.44	1.06 1.24	.67 .83	.76	40.0	51.0			
115	26	10	.22	.46 1.38	.95 1.24	.69 .90	.75	20.0	46.6			
100	26	10	.22	.46 1.30	.80 1.34	.88 .42	.75	30.0	46.6			
77	23	10	-.42	.47 1.97	.07 1.31	.79 .67	.76	60.0	51.0			
103	24	10	-.21	.46 1.25	.69 1.31	.82 .36	.76	40.0	49.0			
41	21	10	-.87	.48 1.28	.72 1.13	.42 .78	.75	50.0	58.4			
46	21	10	-.87	.48 1.28	.72 1.13	.42 .78	.75	50.0	58.4			
230	23	10	-.42	.47 1.27	.73 1.16	.50 .91	.76	30.0	51.0			
251	23	10	-.42	.47 1.22	.63 1.23	.64 .69	.76	50.0	51.0			
119	23	10	-.42	.47 1.22	.62 1.15	.47 .82	.76	30.0	51.0			
229	23	10	-.42	.47 1.22	.62 1.08	.33 .63	.76	80.0	51.0			
131	31	10	1.32	.49 .79	-.33 1.05	.26 .61	.69	50.0	60.4			

Tabel 8. Hasil uji *person misfit order*

Tabel 8 menunjukkan bahwa terdapat 62 responden yang tidak *fit* dengan rasch model. Hasil uji *person misfit order* pada penelitian mengakibatkan adanya *noise* pada hasil pengukuran. Terdapat 5 hal yang dapat menyebabkan responden tergolong dalam kelompok yang tidak *fit* dengan model. Menurut Karabatsos (2003) hal tersebut yakni, 1)

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.

Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

menyontek; 2) kurang teliti; 3) tebakan yang benar; 4) respon yang berlebihan; 5) menjawab dengan acak [35]. Nilai *Infit* MNSQ tertinggi (6,43) dan nilai *Outfit* MNSQ tertinggi (7,28) dimiliki oleh responden nomor 147.

B. Reliabilitas

Hasil uji statistik menunjukkan hasil pengujian reliabilitas pada setiap jenis kelompok aitem. Pengujian ini diperlukan untuk mengetahui kualitas instrumen agar dapat digunakan kapanpun dan dimanapun [36]. Perlu diketahui bahwa terdapat dua jenis reliabilitas pada pengujian reliabilitas menggunakan rasch model, yakni reliabilitas aitem dan reliabilitas *person*. Reliabilitas aitem menunjukkan konsistensi aitem dalam mengukur kemampuan individu, sedangkan reliabilitas *person* menunjukkan konsistensi responden dalam memberikan respon pada setiap aitem [21].

PERSON	251 INPUT		251 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	22.9	10.0		-.45	.53	.99	-.3	1.00
P.SD	2.9	.0		.66	.11	.84	1.6	.90
REAL RMSE	.54	TRUE SD		.37	SEPARATION	.69	PERSON RELIABILITY	.33

ITEM	10 INPUT		10 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	575.4	251.0		.00	.10	1.00	-.2	1.00
P.SD	188.6	.0		1.45	.01	.20	2.4	.22
REAL RMSE	.10	TRUE SD		1.45	SEPARATION	14.37	ITEM RELIABILITY	1.00

Tabel 9. Hasil Pengujian Reliabilitas

Pada hasil uji statistik dalam tabel 9 menunjukkan bahwa nilai reliabilitas aitem = 1,00 ($> 0,94$) dan nilai reliabilitas *person* = 0,33 ($< 0,67$). Nilai reliabilitas aitem menandakan bahwa aitem pada instrumen bersifat sangat reliabel. Kemudian, nilai reliabilitas *person* menandakan bahwa responden memiliki konsistensi jawaban yang lemah. Meski reliabilitas aitem menunjukkan angka yang tinggi, perlu diingat bahwa beberapa aitem instrumen memerlukan pengkajian ulang.

Pada hasil uji statistik juga memuat dua jenis nilai *separation*. Nilai *item separation* berguna untuk menunjukkan rentang tingkat kesulitan aitem secara keseluruhan pada setiap kelompok aitem [25]. Aitem pada instrumen memiliki nilai *item separation* = 14,37 (> 5) yang menandakan bahwa aitem pada instrumen ini memiliki rentang tingkat kesulitan yang sangat bervariasi secara umum. Nilai *person separation* berfungsi untuk mengetahui tingkat variasi kemampuan responden pada setiap kelompok aitem [25]. Hasil uji menunjukkan bahwa responden memiliki nilai *person separation* = 0,69 (< 2). Hal ini secara umum menandakan bahwa kemampuan yang dimiliki responden kurang variatif.

Infit MNSQ menunjukkan informasi statistik *inlier* (respon yang relevan dan konsisten dengan kemampuan individu atau disebut *range* normal dari ekspektasi model). Baik *Infit* MNSQ dan *Outfit* MNSQ memiliki nilai ideal = 1,00 [25]. Hasil analisis menunjukkan bahwa nilai *item Infit* MNSQ = 1,00, *Outfit* MNSQ = 1,00 dan nilai *person Infit* MNSQ = 0,99, *Outfit* MNSQ = 1,00. Dikarenakan kedua hasil pengukuran baik pada aitem dan *person*, mendekati nilai ideal, dapat disimpulkan bahwa secara umum, seluruh aitem instrumen SBBS dan respon responden memenuhi model rasch dengan baik.

2. Pembahasan

Hasil penelitian menunjukkan bahwa instrumen SBBS yang telah diadaptasi memiliki unidimensionalitas dan *local independence* yang cukup baik, namun terdapat beberapa aitem yang tidak *fit* dengan rasch model. Instrumen cukup baik dalam mengukur konstruk yang seharusnya diukur dan cukup tidak terpengaruh oleh varians lain di luar konstruk. Perbandingan dengan hasil uji *Confirmatory Factor Analysis* juga menunjukkan bahwa instrumen ini memiliki validitas konstruk yang baik dengan nilai *loading factor* berkisar antara 0,62-0,89. Selain itu, nilai *Average Variance Extracted* (AVE) pada aspek-aspeknya memiliki nilai *bystander* = 0,586, *defender* = 0,430, dan *probullly* = 0,657. Aspek *defender* memiliki nilai agak rendah ($x < 0,5$) namun masih dapat diterima karena memiliki *composite reliability* yang baik dengan nilai 0,751 [19]. Hal ini menunjukkan bahwa aitem-aitem pada instrumen tersebut memiliki kontribusi kuat dalam mewakili konstruknya dan sejalan dengan hasil pengujian unidimensionalitas. Meski demikian, beberapa aitem penelitian ini tidak memenuhi pengujian *item misfit order* karena terdapat beberapa aitem yang memiliki nilai *Outfit ZSTD* melebihi standar ($x > 2,0$), yakni aitem B2 = -2,38, B3 = -3,17, D2 = 2,91, dan D3 = 4,76. Hal ini menunjukkan bahwa aitem-aitem tersebut memiliki statistik *outlier* yang tinggi dan di luar prediksi rasch model.

Terdapat lima aitem yang memiliki daya beda yang kurang baik, yakni aitem B1, D1, D2, D3, dan D4. Kelima aitem tersebut memiliki nilai *point measure correlation* di bawah standar ($x < 0,4$) dengan nilai pada aitem B1 = 0,17, D1 = 0,14, D2 = 0,22, D3 = 0,15, dan D4 = 0,30. Perlu diperhatikan bahwa hasil uji *point measure correlation* ini telah mengeliminasi semua aitem sikap *defender* sehingga tidak ada satupun aitem yang tersisa pada aspek tersebut.

Hal ini berbeda dengan hasil pengujian *Confirmatory Factor Analysis* yang menunjukkan bahwa aitem-aitem pada instrumen ini memiliki daya yang baik karena memiliki *loading factor* yang mencukupi [19].

Perbedaan daya beda aitem yang didapatkan berasal dari pemilihan diksi yang kurang sesuai untuk usia responden. Hal ini diperkuat dengan temuan penelitian yang menunjukkan bahwa mayoritas responden memiliki kemampuan lebih rendah dibanding tingkat kesulitan aitem, yang mana hal ini mengindikasikan bahwa mayoritas responden kurang dapat memahami aitem-aitem pada instrumen penelitian. Peninjauan terhadap butir respon pada *empirical item-category function* juga menunjukkan bahwa butir respon memiliki jarak yang cukup dekat yang mengindikasikan bahwa butir respon kurang dapat membedakan kemampuan responden dengan baik. Selain itu, responden juga menunjukkan kebingungan terhadap butir respon pada aitem B1 dan D3 dikarenakan aitem-aitem tersebut memiliki susunan skor yang terbalik pada hasil uji *empirical item-category function*. Selain itu, hasil uji *person misfit order* menunjukkan bahwa terdapat 62 responden yang tidak *fit* dengan rasch model. Responden yang tidak *fit* mengakibatkan *noise* sehingga instrumen mengukur hal-hal lain yang seharusnya diukur.

Meskipun hasil penelitian menunjukkan bahwa beberapa aitem pada instrumen penelitian bersifat kurang valid dan memerlukan pengkajian ulang, aitem pada instrumen yang digunakan dalam penelitian ini memiliki tingkat kesulitan aitem yang baik. Hal ini didasarkan pada nilai *item separation* = 14,37 (> 5) yang menunjukkan bahwa instrumen penelitian memiliki variasi tingkat kesulitan yang tinggi. Selain itu, instrumen juga bersifat sangat reliabel dengan nilai reliabilitas aitem = 1,00 ($> 0,94$). Ini menunjukkan bahwa instrumen sangat konsisten dalam mengukur kemampuan responden.

Sebagaimana yang telah dibahas sebelumnya, beberapa aitem pada instrumen ini memiliki validitas yang kurang baik dikarenakan pemilihan diksi kurang sesuai untuk usia responden dan adanya *noise* dari berasal dari responden yang tidak *fit* dengan rasch model. Oleh karena itu, dibutuhkan pengkajian ulang terhadap aitem-aitem pada instrumen agar dapat memperbaiki masalah ini. Selain itu, peneliti menemukan bahwa terdapat kecurangan dan kebingungan pada diri responden ketika proses pengumpulan data. Pemberian pengawasan dan pemahaman yang lebih baik ketika proses pengumpulan data juga dapat menjadi solusi guna meminimalisir kecurangan dan ketidakpahaman responden terhadap aitem-aitem pada instrumen.

IV. KESIMPULAN

Pengembangan instrumen penelitian dengan metode adaptasi terhadap instrumen *School Bystander Behaviour Scale* (SBBS) untuk mengukur kecenderungan sikap siswa *bystander* dalam *bullying* di Indonesia dengan metode analisis rasch model menunjukkan bahwa beberapa aitem pada instrumen SBBS yang telah diadaptasi kurang dapat menunjukkan performa yang baik dikarenakan memiliki validitas yang kurang baik. Meski demikian, secara keseluruhan, instrumen memiliki validitas yang cukup karena bersifat unidimensional dan cukup memiliki *local independence* sehingga instrumen dapat menukur konstruk yang dimaksud dan cukup tidak terpengaruh varians lain di luar konstruk. Hasil uji reliabilitas juga menunjukkan bahwa reliabilitas yang dimiliki instrumen tergolong sangat tinggi sehingga instrumen bersifat sangat reliabel serta memiliki aitem dengan tingkat kesulitan yang sangat bervariasi. Oleh karena itu, direkomendasikan pengkajian lebih lanjut terhadap beberapa aitem agar instrumen ini dapat memberikan sifat psikometrik dan performa yang lebih baik. Pemberian pemahaman dan pengawasan yang lebih baik kepada responden ketika proses pengumpulan data juga dapat menjadi solusi guna meminimalisir ketidakpahaman dan kecurangan yang mungkin dilakukan oleh responden yang mana hal ini akan berkontribusi terhadap berkurangnya *noise* pada temuan penelitian. Selain itu, direkomendasikan untuk melakukan penelitian *Confirmatory Factor Analysis* terhadap instrumen yang telah diadaptasi dalam Bahasa Indonesia ini. Penelitian dengan populasi yang berbeda juga disarankan karena nilai dan norma yang berada di suatu daerah dapat berbeda dengan daerah lainnya sehingga dapat mempengaruhi *bystander effect*.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih sepenuh hati kepada instansi SMP Negeri 2 Pungging, Mojokerto, atas izin yang diberikan untuk melaksanakan penelitian di sekolah ini dan seluruh siswa SMP Negeri 2 Pungging, Mojokerto, yang bersedia menjadi responden penelitian dan berpartisipasi aktif dalam pengumpulan data.

REFERENSI

- [1] P. Fischer *et al.*, “The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies.,” *Psychol. Bull.*, vol. 137, no. 4, Art. no. 4, 2011.
- [2] J. Hudson and A. Bruckman, “The Bystander Effect: A Lens for Understanding Patterns of Participation,” *J. Learn. Sci. - J LEARN SCI*, vol. 13, pp. 165–195, Apr. 2004.

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.

Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

- [3] A. Halimah, A. Khumas, and K. Zainuddin, "Persepsi pada Bystander terhadap Intensitas Bullying pada Siswa SMP," *J. Psikol.*, vol. 42, p. 129, Aug. 2015.
- [4] E. Haru, "Perilaku Bullying Di Kalangan Pelajar," *J. Altern. Wacana Ilm. Interkultural*, vol. 11, no. 2, Art. no. 2, 2022.
- [5] A. Prasetyo, "Bullying di Sekolah dan Dampaknya bagi Masa Depan Anak," *El-Tarawwi*, vol. 4, pp. 19–26, Jan. 2014.
- [6] N. Karisma, A. Rofiah, S. Afifah, and Y. Manik, "Kesehatan Mental Remaja dan Tren Bunuh Diri: Peran Masyarakat Mengatasi Kasus Bullying di Indonesia," *Edu Cendikia J. Ilm. Kependidikan*, vol. 3, pp. 560–567, Jan. 2024.
- [7] W. D. Lestari, T. M. Anggriana, and B. D. Pratama, "Pengaruh Empati Dan Bystander Effect Terhadap Perilaku Prososial Siswa SMP," *Pros. Semin. Nas. Bimbing. Dan Konseling*, vol. 4, no. 1, Art. no. 1, May 2022.
- [8] M. B. A. Tumon, "Studi Deskriptif Perilaku Bullying Pada Remaja," *Calyptra*, vol. 3, no. 1, Art. no. 1, Sept. 2014.
- [9] R. Thornberg and T. Jungert, "Bystander behavior in bullying situations: Basic moral sensitivity, moral disengagement and defender self-efficacy," *J. Adolesc.*, vol. 36, pp. 475–483, Mar. 2013.
- [10] R. Thornberg, T. Pozzoli, and G. Gini, "Defending or Remaining Passive as a Bystander of School Bullying in Sweden: The Role of Moral Disengagement and Antibullying Class Norms," *J. Interpers. Violence*, vol. 37, pp. 1–24, Aug. 2021.
- [11] C. Salmivalli, "Bullying and the peer group: A review," *Aggress. Violent Behav.*, vol. 15, pp. 112–120, Mar. 2010.
- [12] S. Padgett and C. Notar, "Bystanders are the Key to Stopping Bullying," *Univers. J. Educ. Res.*, vol. 1, pp. 33–41, Aug. 2013.
- [13] N. M. D. Sulistiowati, I. G. A. N. F. Wulansari, K. E. Swedarma, A. P. Purnama, and N. P. Kresnayanti, "Gambaran Perilaku Bullying Dan Perilaku Mencari Bantuan Remaja SMP Di Kota Denpasar," *J. Ilmu Keperawatan Jiwa*, vol. 5, no. 1, Art. no. 1, Jan. 2022.
- [14] S. F. D. Maharani and I. S. Borualogo, "Hubungan antara Iklim Sekolah dan Subjective Well-Being Siswa SMP Korban Perundungan Siber di Kota Bandung," *Bdg. Conf. Ser. Psychol. Sci.*, vol. 2, no. 2, pp. 622–631, July 2022.
- [15] R. A. Sestiani and A. Muhib, "Pentingnya Dukungan Sosial Terhadap Kepercayaan Diri Penyintas Bullying: Literature Review," *TEMATIK*, vol. 2, no. 1, Art. no. 1, Jan. 2022.
- [16] A. Nickerson, A. Aloe, J. Livingston, and T. Feeley, "Measurement of the bystander intervention model for bullying and sexual harassment," *J. Adolesc.*, vol. 37, pp. 391–400, June 2014.
- [17] M. Demaray, K. Summers, L. Jenkins, and L. Davidson Becker, "Bullying Participant Behaviors Questionnaire (BPBQ) : Establishing a Reliable and Valid Measure," *J. Sch. Violence*, vol. 15, Dec. 2014.
- [18] F. Goossens and T. Olthof, "The New Participant Role Scales: A comparison between various criteria for assigning roles and indications for their validity," *Sage J.*, vol. 26, pp. 146–151, Apr. 2017.
- [19] Z. Suárez-García, R. Thornberg, and D. Álvarez-García, "Validation of a Scale for Assessing Bystander Responses in Bullying," *Psicothema*, vol. 33, pp. 623–630, Oct. 2021.
- [20] A. Costa, M. Vives, and J. Corey, "On Language Processing Shaping Decision Making," *Curr. Dir. Psychol. Sci.*, vol. 26, pp. 146–151, Apr. 2017.
- [21] Y. E. Suryani, "Aplikasi Rasch Model dalam Mengevaluasi Intelligenz Structure Test (IST)," *Psikohumaniora J. Penelit. Psikol.*, vol. 3, no. 1, Art. no. 1, Aug. 2018.
- [22] D. E. Beaton, C. Bombardier, F. Guillemin, and M. B. Ferraz, "Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures;," *Spine*, vol. 25, no. 24, Art. no. 24, Dec. 2000.
- [23] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif Dan R & D*. Bandung: Alfabeta, 2017. Accessed: Sept. 04, 2024.
- [24] I. Machali, *Metode Penelitian Kuantitatif (Panduan Praktis Merencanakan, Melaksanakan, dan Analisis dalam Penelitian Kuantitatif)*. Yogyakarta: Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri (UIN) Sunan Kalijaga Yogyakarta, 2021.
- [25] B. Sumintono and W. Widhiarso, *Model Rasch untuk Penelitian Sosial Kuantitatif*. Surabaya, 2014.
- [26] S. Golia and E. Brentari, "Unidimensionality in the Rasch model: how to detect and interpret," *Statistica*, vol. 67, no. 3, pp. 253–261, Sept. 2007.
- [27] F. M. Hamzah, M. N. A. Rashid, M. N. A. R. Rahman, and Md. M. Rasul, "Rasch Model Analysis for Evaluating Validity and Reliability of Authentic Learning Instruments in Malaysian Polytechnics," *Int. J. Innov.*, vol. 9, no. 7, 2019.

- [28] G. Hamdu, F. N. Fuadi, A. Yulianto, and Y. S. Akhirani, "Items Quality Analysis Using Rasch Model To Measure Elementary School Students' Critical Thinking Skill On Stem Learning," *JPI J. Pendidik. Indones.*, vol. 9, no. 1, Art. no. 1, Mar. 2020.
- [29] K. P. Prayoga, D. Suryana, M. Supriatna, and N. Budiman, "Penggunaan Rasch Model Untuk Menganalisis Konstruk Instrumen Kontrol Diri Pada Siswa Sekolah Menengah," *G-Couns J. Bimbing. Dan Konseling*, vol. 9, no. 1, pp. 367–381, Aug. 2024.
- [30] Y. Handayani, R. Rahmawati, and W. Widiasih, "Using Rasch Model to Analyze Reliability and Validity of Concept Mastery Test on Electricity and Magnetism Topic," *JIPF J. Ilmu Pendidik. Fis.*, vol. 8, no. 2, pp. 226–239, May 2023.
- [31] S. Hayati, S. W. Hidayanti, and A. Rizki, "Analisis Rasch Model: Mengukur Kualitas Butir Soal Tes Tashrif Lughawi," *Intelektualita*, vol. 13, no. 1, Art. no. 1, July 2024.
- [32] I. F. Apriani, R. W. Muhamram, I. Rachmawati, L. Susilawati, and G. Islamati, "Kemampuan Perkalian dan Pembagian Bilangan Cacah Menggunakan RASCH Model untuk Siswa Kelas IV SD," *Dwija Cendekia J. Ris. Pedagog.*, vol. 7, no. 2, Aug. 2023.
- [33] I. Rivaldo, H. Sutrisno, and A. C. Manik, "The Use of The Rasch Model to Develop Students' Conception of Chemistry Learning Instruments During the Covid-19 Pandemic," *J. Pendidik. Kim. Indones.*, vol. 6, no. 1, Art. no. 1, May 2022.
- [34] Y. A. Rahman, L. H. Rentina, and U. R. Dhini, "Person Fit Analysis For Assessing Academic Writing Performance Using Rasch Model," *J. Pendidik. Glas.*, vol. 7, no. 2, Art. no. 2, July 2023.
- [35] G. Karabatsos, "Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics," *Appl. Meas. Educ.*, vol. 16, no. 4, pp. 277–298, 2003.
- [36] A. Anggraini and M. Muntazhimah, "Pengembangan Instrumen Kemampuan Berpikir Reflektif Matematis Siswa Madrasah Aliyah," *Aksioma J. Program Studi Pendidik. Mat.*, vol. 10, no. 4, Art. no. 4, Dec. 2021.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.