The 21st-century skills scales: many facet Rasch measurements

Ma'rifatin Indah Kholili[1], Adi Dewantoro[1], Naharus Surur[1], Novita Tri Hapsari[2]
[1]Guidance and Counseling Department, Faculty of Teacher Training and Education, Sebelas Maret University, Surakarta, Indonesia
[2]Guidance and Counseling Department, Faculty of Education and Psychology, Yogyakarta State University, Yogyakarta, Indonesia

Article Info

ABSTRACT

This study aims to develop a 21st century skills scale for prospective school counselor. The method used is the research and development. The data were analyzed using content analysis/expert judgment with many-facet Rasch model (MFRM). Three facets were employed: scale, item, and measurement criteria. The initial version of the scale comprises 129 items measuring four aspects: way of thinking, way of working, tools for working, and living in the world. However, 78 items were deemed valid following the expert judgment analysis result. The inter-rater agreement opportunity shows raters' similarity score of 10,836, implying that experts in this study tend to give similar assessments. This research results in a 21st century skills scale for prospective school counselors with a total of 78 valid items. The purpose of this study was to develop a 21st century skills scale for prospective school counselors that had been fulfilled, the instrument had been tested for validity and reliability based on expert judgment. The current research recommends that further research be carried out to develop a 21st century skills scale instrument for prospective educators in other scientific fields.

Corresponding Author:

Ma'rifatin Indah Kholili
Guidance and Counseling Department, Faculty of Teacher Training and Education,
Sebelas Maret University
Jebres, Surakarta, Central Java 57126, Indonesia
Email: marifatin.ink@staff.uns.ac.id

1. INTRODUCTION

Education is an issue associated with all individuals, either personally or professionally, at national and global levels [1]. There need to be skilled in carrying out education, especially to face the 21st century. There is no globally accepted definition for 21st century skills. However, it refers to a set of skills students expected to possess to ensure a successful future career. It is also viewed as how students understand these skills and their attitudes toward the world of work. Education department students are expected to possess skills to ensure their future success [2], [3]. The 21st century skills have been defined in various international organizations and projects, such as assessment and teaching of 21st century skills (AT21CS), partnership 21 (P21) century skills, organization for economic co-operation and development (OECD) competency definition and selection, and European union's key competences for lifelong learning [2].

The AT21CS consortium administer the 21st-century skill, knowledge, attitude, value, and ethics into four categories: i) ways of thinking, creativity and innovation, critical thinking, problem-solving, decision-making, and learning to learn; ii) ways of working, communication and teamwork; iii) tools for working, general knowledge and ICT literacy; and iv) living in the world, citizenship, life and career, personal and social responsibilities, including cultural awareness and competence [4]. The P21 century skills proposes a collective, integrated learning visions to help practitioners integrate these skills into the academic learning contexts, known as the 21st-century learning framework: i) learning skills and innovation, creativity

and innovation, critical thinking and problem-solving skills, communication and collaboration; ii) Information, media, and technology skills, information media, and ICT literacies; iii) life skills and career, flexibility and adaptability, initiative and self-directed social and cross-cultural skills, productivity and accountability, leadership and responsibility P21 century skills [5].

Aspects and indicators of the 21st century skills in this study were derived from concepts proposed by P21 century skill and AT21CS [4]. The term 21st century skill is multidisciplinary in nature, and is relevant to various aspects of contemporary life in this increasingly complex world. This skill mostly involves skills and understanding as its aspects, while some literature argue that it emphasizes inclination, such as curiosity, creativity, and collaboration, which are basically not a skill. Other literature also emphasizes the technological aspect, while others believe it emphasizes attitude and values. Despite this difference, most literature focus on thinking, learning, and complex communication skill that demands a learning process than merely memorizing [6]. Skills are acquired, observable, and implemented, and easily recognized in almost all settings [7].

Skills for prospective school counselor in 21st century are closely linked to innovation and technologies. In order to effectively involve Generation Z in the learning process, the education system the needs to fulfill needs for these skills [8]. Preparing 21st century students to adapt to professional and social values is a complicated task. Education department students are demanded to keep growing to cope with challenges in society, including technology development and changes and professional life skills [9]. Current globalization, technology, migration, international competition, market and environmental changes, and international politics increase the urgency of 21st century skills and knowledge for students to succeed [10], [11]. School counselors have an important role in the development of students. Professional school counselor (PSC) has shown their role in addressing many students' problems. School counselor educators need to get involved in helping prepare the skills of future counselor candidates by providing practical experience [12]. Prospective school counselors are students who take formal education to have the skills to become professional school counselors.

The importance of training for aspiring counselors will assist them in utilizing the skills needed in school Prospective counselors also need these skills to help learners face the challenges of the 21st century. School counselors must be able to deal with developmental issues, being the role and responsibility of counselors to facilitate learners' success [13], [14]. As illustrated by the school counselor's job description, prospective school counselors have a dual role to be able to use a variety of knowledge and skills. In recent years, especially during pandemics, many people sought for mental health services. It is being even more currently. This situation makes counselor skills in 21st century skills centuries should be improved to capable to give appropriate service [15]. These skills are useful for enriching the counselor's ability in the process of providing services for students in schools.

The 21st-century skills a student need include problem-solving, communication, teamwork, critical thinking, creative thinking, and ICT skills. To provide students with such abundant skills, 21st century teachers should possess various skills and competencies, such as continuous development, empathy, effective communication, problem-solving, personality demonstration, and guidance [11]. The 21st-century teachers' skills and competence will likely facilitate students to achieve optimal development. They also play a pivotal role in establishing a positive learning environment, either in academic, emotional, and social aspects. Counselors in the 21st century are instrumental in shaping a positive environment in learning both academically, emotionally, and socially. School counselors must pay attention to students' mental and social-emotional needs [16].

Skills are not easy to measure, but an instrument development process can increase understanding of how skills are assessed at any given moment [17]. As a place to produce qualified teacher candidates, universities must begin to conduct innovative assessments of the competencies and skills of prospective teachers [18]. In fact, the 21st-century skills scale for prospective school counselor students has yet to be available. Therefore, it is necessary to develop a scale to measure students' 21st century skills. Measurement can be thought of as a process involving three components -an object of measurement, a set of numbers, and a system of rules- that assign a number to the magnitude of the measured variable [19]. Rahman *et al.* [20] have been exploring the instrument to measure counselor competency but limited to the Islamic counselor only.

Seeing the seriousness of the problems, academicians need efforts to be able to produce the instruments needed following developments. This study aims to produce a 21st-century skill scale instrument for prospective school counselor students whom experts validate. Through the development of a systematic scale, the level of acceptability of the instrument can be known. This research can bridge and fulfill the availability of 21st century skills instruments for prospective school counselors which are not yet available. This instrument can be useful for guidance and counseling lecturers and students so they can measure the level of their 21st century skills.

2. RESEARCH METHOD

2.1. Research design

The method used is scale research and development, which refers to the DeVellis model. There are 8 stages: determine the aims of measure, generate an item pool, determine the format for measurement, have initial item pool reviewed by experts, consider inclusion of validation items, administer items to a development sample, evaluate the items, optimize scale length [21]. The scale development method follows the procedure presented by DeVellis [21]:

− Determine the aims of the measure: The first step that must be done is to determine what we want to measure. What needs to be considered in determining the goal is that the measurements made must be based on the substantive theory.

− Generate an item pool: In the second step, the developer is ready to create the instrument. Things to note that the items written must reflect the purpose of measurement or according to the construct you want to measure, the number of items, and the use of positive and negative items.

− Determine the format for measurement: This third step is to determine the scale format to be used.

− Have initial item pool reviewed by experts: The next step in this process is to have a group of knowledgeable people review or evaluate the set of items in the content area. This review was conducted with several objectives that maximize the validity of the content of the scale.

− Consider the inclusion of validation items: The sixth step is the essence of scale development, namely the collection of items developed has been tested for validation.

− Administer items to a development sample: Having decided on construct-related items and which validity to include in the questionnaire, the developer must manage them. Item management considers the number of subjects in the sample.

− Evaluate the items: Once the initial set of items has been developed, researched, and given to a sufficiently large and representative sample, the next stage is to evaluate the performance of each item so that the corresponding items can be identified to form a scale.

− Optimize scale length: At this last stage, the developer has a collection of items demonstrating acceptable reliability.

2.2. Participants

This research was conducted from August 2021-January 2022. The research site was centered in the city of Surakarta, Central Java Province, but for participants involved experts from East Java, Central Java, and West Java, Indonesia. The expert validity test was carried out by involving eight guidance and counseling experts. The process of selecting experts is determined based on the following criteria: minimum doctoral education in the field of guidance and counseling, has a minimum ten years work experience, competent in scale development, experts are selected from different universities.

2.3. Data collection instruments

Research data collection using expert assessment questionnaires. Expert answer choices for each item use a Likert scale. The Likert scale is one of the most fundamental psychometric tools and is common for social and educational science studies [22]–[24]. The alternative responses include 1 (strongly irrelevant), 2 (less relevant), 3 (relevant) and 4 (strongly relevant). Each statement item is assessed from four ranges of assessment figures based on three criteria: usability, accuracy, and feasibility. An example of an expert questionnaire instrument is presented in Table 1.

Table 1. Expert questionnaire instruments

Indicators	Descriptors	Item	Usability	Accuracy	Feasibility
Ways of thinking	Creativity and innovation	I love thinking about innovations for my activities			
		I am looking for the most practical way to get the job			
		My mind is filled with interesting ideas			

Each item was assessed in a 1-4 score range based on three criteria: usability, accuracy, and feasibility. The first criterion, i.e., usability, refers to the usefulness of each item in the 21st century skill. An item was deemed useful when it is relevant to the conceptual and operational definitions of 21st century skills, construct-clear (complying with the indicators and descriptors), meaningful and valuable for scale development. The second criterion, i.e., feasibility, highlights the procedure of item development.

The indicators of feasibility include compliance with the instrument development procedure, relevance to the current needs, thoroughness in developing each item, and use of relevant references. Lastly, the third criterion, i.e., accuracy, refers to the language use of the developed item. An accurate item is indicated by its easy-to-understand language use, effective sentence structure and does not cause ambiguity. The validation of this assessment questionnaire only consulted one expert to be given qualitative input. The result was obtained input that in each item, the developed statement must be filled out by an expert based on all three criteria of usability, accuracy, and feasibility.

2.4. Expert judgment analysis

There are several types of validity: face validity, content validity, construct validity, criterion validity and reliability [25], [26]. This study focused on content validity/expert judgment. It represents the degree to which an instrument (e.g., a questionnaire) represents all the desired aspects of a theoretical construct (as judged by experts) [27]. The content validity was performed by involving a content expert panel, the Delphi method, or representative groups. Content validity refers to the scale ability to measure the item to be measured accurately. In social science, content validity could be developed through several methods, including an expert panel, literature review, and Delphi approach [28].

In this study, experts were recruited to test the scale validity. The data were analyzed using many-facet Rasch model (MFRM). Many researchers prefer Rasch model to guide the scale construction to additive model preference like Cronbach's alpha or factor analysis. Rasch model, as Guttman conveys, set an item to measure a construct and establish an orderly relationship [29]. MFRM was developed by Linacre to address the variability exhibited by several raters [30], [31]. Its advantages lie in its ability to model the raters according to their responses and define the scale for raters personally, meaning that the model could depict the raters identically [31].

MFRM is a probability-based logistic measurement. Thus, the logarithm capable of providing information about the scale accuracy, reliability, and validity were means-square (MNSQ) outfit, a standardized outfit (ZSTD), and point measure correlation (Pt.MeaCor) values [32]. The analysis result showed the outfit and infit value of test items. Infit (inlier sensitive or information weighted fit) refers to response pattern sensitivity to the target item on the respondent, or vice versa, while outfit (outlier sensitive fit) represent the response pattern sensitivity to items with certain difficulty on respondents, or vice versa [33]. MFRM could measure interactions between facets, which may indicate unexpected responses or bias in assessment process [34], [35]. The present study used three facets: i) scale (eight experts); ii) item (n=128); and iii) measurement criteria (3 criteria).

3. RESULTS AND DISCUSSION
3.1. The 21st century skill scale development process

Scale development is a process of designing a reliable and valid construct to measure an attribute to be measured. There are several principles in the scale development process, including approaches to scale creation, construct definition, purposes of created scale, principles of item writing, scale validation research design, scale psychometric properties, scale revision [36]. The first step of scale development is to determine goals. The purpose of scale development is in line with the purpose of this study, which is to produce a 21st century skill-measuring tool for prospective school counselor students. The theoretical basis used to develop 21st century skills is derived from the concepts of P21 century skills and AT21CS.

The second step is to compile the statement item. The development process of this scale only uses favorable items without any unfavorable items. Items are developed based on four aspects: ways of thinking/learning and innovation skills (37 items); ways of working/life and career skills (39 items); tools for working/information, media, and technology skills (21 items); and living in the world (32 items). From each aspect, it is lowered into an indicator and then compiled the statement items. There are 129 items in total, and the more complete 21st century skill instrument grid is in Table 2.

The third step is to determine the scale format. This developed instrument uses a Likert scale format consisting of very suitable, appropriate, less suitable, and unsuitable options. The fourth step of scale development is expert review/assessment. This step is done to test the validity of the contents of the instrument. The experts involved are eight guidance and counseling experts (student lecturers who are prospective school counselors) from various universities. Expert assessment consists of qualitative input and quantitative assessment. The results of the quantitative assessment of expert validators in detail are presented and analyzed in the next section.

Table 2. The 21st century skill instrument grid

Variable	Aspects		Indicator	No. of items
21St-century skills	Ways of thinking (learning and innovation skills)	a.	Creativity and innovation	9
		b.	Critical thinking	12
		c.	Problem solving	4
		d.	Decision making	6
		e.	Learning to learn or metacognition	6
	Ways of working (life and career skills)	a.	Flexibility and adaptability	6
		b.	Initiative and self-direction	2
		c.	Social and cross-cultural skills	7
		d.	Productivity and accountability	6
		e.	Leadership and responsibility	10
		f.	Communication and teamwork	8
	Tools for working (information, media and technology skills), Living in the world.	a.	Information literacy	7
		b.	Media literacy	8
		c.	General knowledge and ICT literacy	6
		a.	Citizenship	8
		b.	Personal and social responsibility	8
		c.	Cultural awareness and competence	8
		d.	Spiritual-religious	8
	Total			129

3.2. Expert validator assessment result
3.2.1. Multirater validation

Figure 1 presents the multirater validation roadmap according to the many-facet Rasch model (MFRM). Figure 1 displays the general result of expert validation test. Figure 1 presents five columns. Column 1 presents the assessment result with a score range between +2 and -5, known as logit values. Logit value of 0 [29] is the minimum criteria of an item deemed to have good quality by the expert. Hence, the positive value (higher than logit of 0) means that the item was categorized as a good item by the expert panel, while the negative logit value (less than 0) means that the item was poor, according to the experts in this study. Column 2 presents items assessed by experts, showing that the lowest scoring distribution was close to logit of -2, while the highest was close to logit of +2. Column 3 presents the expert assessment criteria, including accuracy, feasibility, and usability. Column 4 displays the validators' initials and shows that the eight validators in this study were in negative logit value (-2 to -5). This indicates that eight validators in this study were less stringent in assessing each item. Column 5 presents the assessment's 4-point scale used to assess each item.
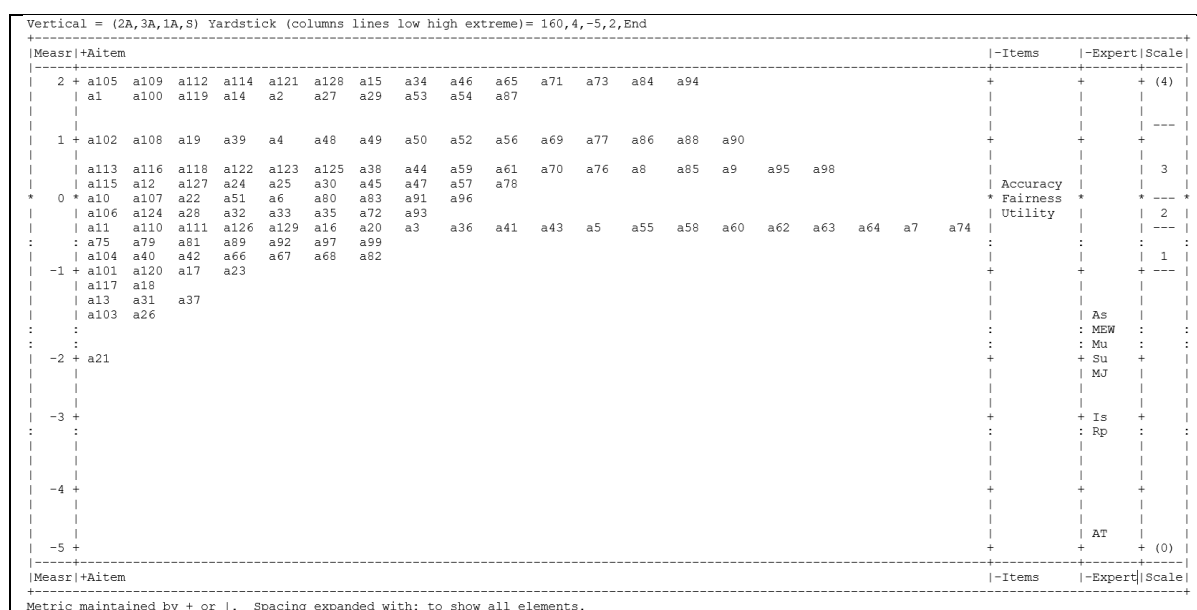
```
Vertical = (2A,3A,1A,S) Yardstick (columns lines low high extreme)= 160,4,-5,2,End
+------------------------------------------------------------------------------+
|Measr|+Aitem                                                    |-Items  |-Expert|Scale|
|-----+----------+----------------------------------------------+--------+-------+-----+
|  2 + a105 a109 a112 a114 a121 a128 a15 a34 a46 a65 a71 a73 a84 a94    +       +     + (4) |
|   | a1   a100 a119 a14 a2 a27 a29 a53 a54 a87                         |       |     |     |
|   |  |                                                                |       |     |     |
|   |  |                                                                |       |     | --- |
|  1 + a102 a108 a19 a39 a4 a48 a49 a50 a52 a56 a69 a77 a86 a88 a90     +       +     +     |
|   |  |                                                                |       |     | 3   |
|   | a113 a116 a118 a122 a123 a125 a38 a44 a59 a61 a70 a76 a8 a85 a9 a95 a98   Accuracy |     |
|   | a115 a12 a127 a24 a25 a30 a45 a47 a57 a78                          |       |     |     |
|*  0 * a10 a107 a22 a51 a6 a80 a83 a91 a96                          * Fairness *     * --- *
|   | a106 a124 a28 a32 a33 a35 a72 a93                                 | Utility |   | 2   |
|   | a11 a110 a111 a126 a129 a16 a20 a3 a36 a41 a43 a5 a55 a58 a60 a62 a63 a64 a7 a74 |   | --- |
|   : a75 a79 a81 a89 a92 a97 a99                                      :       :     |     |
|   | a104 a40 a42 a66 a67 a68 a82                                     |       |     | 1   |
| -1 + a101 a120 a17 a23                                                +       +     + --- |
|   | a117 a18                                                          |       |     |     |
|   | a13 a31 a37                                                       |       |     |     |
|   | a103 a26                                                          |       | As  |     |
|   :                                                                   :       : MEW |     |
|   :  :                                                                :       : Mu  :   : |
| -2 + a21                                                              +       + Su  +     |
|   |  |                                                                |       | MJ  |     |
|   |  |                                                                |       |     |     |
|   |  |                                                                |       |     |     |
| -3 +                                                                  +       + Is  +     |
|   |  |                                                                :       : Rp  :   : |
|   |  |                                                                |       |     |     |
|   |  |                                                                |       |     |     |
| -4 +                                                                  +       +     +     |
|   |  |                                                                |       |     |     |
|   |  |                                                                |       | AT  |     |
| -5 +                                                                  +       +     + (0) |
|-----+----------------------------------------------------------+--------+-------+-----+
|Measr|+Aitem                                                    |-Items  |-Expert|Scale|
+------------------------------------------------------------------------------+
Metric maintained by + or |.  Spacing expanded with: to show all elements.
```

Figure 1. Road map MFRM–validation of expert judgment

3.2.2. Item difficulty distribution

The item difficulty distribution is presented in column 2, as displayed in Figure 1. It presents the validators' responses to the scale item. The column item presents the distribution of the three criteria: accuracy, feasibility, and usability, meaning that each developed item was assessed by the experts based on these three criteria. Items with lowest/negative scores have negative logit values, indicating that the item is among the easiest item to meet the criteria, while those with the highest positive scores have positive logit value, meaning that the item is among the most difficult to meet the criteria. Items with the same logit values have the same level of difficulty/ease. As shown in Figure 1, items a105, a109, a112, a114, a121, a128, a15, a34, a46, a65, a71, a73, a84, a94, a1, a100, a119, a14, a2, a27, a29, a53, a54, a87 exhibited highest logit value, i.e., +2, indicating that these items were the most difficult to meet among other items assessed by validators. Meanwhile, item a21 exhibits the lowest logit value (i.e., -2), indicating that it is the easiest item to meet, according to the eight expert validators.

3.2.3. Validator ease level

Column 4 in Figure 1 shows that eight validators in this study were in negative logit values (-2 to -5), meaning that eight validators in this study tend to be "generous" in assessing each item. The list of validators, sorted from the most severe to the most lenient, is presented in Figure 2. The figure displays the validators' ease level, sorting from the most stringent to the most generous in performing the assessment. Validator AS was considered the most stringent, meaning that AS viewed the items were poor based on the three criteria. AT was found to be the most lenient validator in assessing the items based on the three criteria.

Figure 2 also allows us to see whether or not these validators work independently. The inter-rater agreement opportunity shows raters' similarity score of 10,836, implying that validators in this study tend to give similar judgments. The most similar scoring was 7,555 (69.7%). Validators in this study exhibited a 30.3% variability, implying that they did not work independently. In this regard, higher variability would be better. This could be accounted for by the excessively large number of items, conflict of interest, and bias in instrument development and assessment. In addition to providing the difficulty in assessment, MFRM analysis also shows the validators' assessment consistency, as presented in Figure 3. This figure displays validators' consistency. It shows that validator MU was the most inconsistent validator, while AT was the most consistent validator in this study. This result could be used as a reference for the next process by contacting experts who exhibited consistent assessment.

```
+-----------------------------------------------------------------------------------------------------+
| Obsvd  Obsvd  Obsvd  Fair(M)|   -    Model | Infit     Outfit   |Estim.| Correlation | Exact Agree. |              |
| Score  Count  Average Average|Measure  S.E. | MnSq ZStd MnSq ZStd |Discrm| PtMea PtExp | Obs %  Exp % | N Expert BK  |
|-----------------------------------------------------------------------------------------------------|
|  1235    345   3.58   3.69  | -1.71   .09  | .92  -.6 1.01   .1 |  .96 |  .37   .41  | 60.8  66.4  | 3 As         |
|  1242    345   3.60   3.70  | -1.76   .09  | .99   .0  .70 -2.4 | 1.22 |  .46   .41  | 65.6  67.0  | 4 Mu         |
|  1247    345   3.61   3.71  | -1.80   .09  | .93  -.4 1.04   .3 |  .94 |  .41   .40  | 65.6  67.4  | 1 MEW        |
|  1275    345   3.70   3.77  | -2.06   .10  | 1.33 2.0 1.19  1.2 |  .92 |  .36   .37  | 69.3  69.9  | 7 Su         |
|  1293    345   3.75   3.81  | -2.26   .11  | 1.19 1.1  .86  -.8 | 1.02 |  .40   .34  | 71.9  71.5  | 2 MJ         |
|  1336    345   3.87   3.90  | -3.02   .16  | .85  -.8  .92  -.3 | 1.03 |  .29   .26  | 75.2  75.3  | 5 Is         |
|  1336    345   3.87   3.90  | -3.02   .16  | .99   .0 1.22  1.0 |  .93 |  .18   .26  | 72.5  75.3  | 6 Rp         |
|  1371    345   3.97   3.98  | -4.71   .34  | 1.03  .1 1.39   .8 |  .95 |  .06   .12  | 76.9  77.8  | 8 AT         |
|-----------------------------------------------------------------------------------------------------|
|  1291.9 345.0  3.74   3.81  | -2.54   .14  | 1.03  .2 1.04   .0 |      |  .32        |             | Mean (Count: 8)     |
|    47.7   .0    .14    .10  |   .96   .08  | .15   .9  .21  1.1 |      |  .13        |             | S.D. (Population)   |
|    51.0   .0    .15    .11  |  1.02   .08  | .16  1.0  .22  1.2 |      |  .13        |             | S.D. (Sample)       |
+-----------------------------------------------------------------------------------------------------+
Model, Populn: RMSE .16  Adj (True) S.D. .94  Separation 5.80  Strata 8.07  Reliability (not inter-rater) .97
Model, Sample: RMSE .16  Adj (True) S.D. 1.01  Separation 6.21  Strata 8.62  Reliability (not inter-rater) .97
Model, Fixed (all same) chi-squared: 174.7  d.f.: 7  significance (probability): .00
Model,  Random (normal) chi-squared:  6.5  d.f.: 6  significance (probability): .37
Inter-Rater agreement opportunities: 10836  Exact agreements: 7555 = 69.7%  Expected:  7728.4 =  71.3%
```

Figure 2. Validator ease level

```
+-----------------------------------------------------------------------------------------------------+
| Obsvd  Obsvd  Obsvd  Fair(M)|   -    Model | Infit     Outfit   |Estim.| Correlation | Exact Agree. |              |
| Score  Count  Average Average|Measure  S.E. | MnSq ZStd MnSq ZStd |Discrm| PtMea PtExp | Obs %  Exp % | N Expert BK  |
|-----------------------------------------------------------------------------------------------------|
|  1242    345   3.60   3.70  | -1.76   .09  | .99   .0  .70 -2.4 | 1.22 |  .46   .41  | 65.6  67.0  | 4 Mu         |
|  1336    345   3.87   3.90  | -3.02   .16  | .85  -.8  .92  -.3 | 1.03 |  .29   .26  | 75.2  75.3  | 5 Is         |
|  1235    345   3.58   3.69  | -1.71   .09  | .92  -.6 1.01   .1 |  .96 |  .37   .41  | 60.8  66.4  | 3 As         |
|  1247    345   3.61   3.71  | -1.80   .09  | .93  -.4 1.04   .3 |  .94 |  .41   .40  | 65.6  67.4  | 1 MEW        |
|  1293    345   3.75   3.81  | -2.26   .11  | 1.19 1.1  .86  -.8 | 1.02 |  .40   .34  | 71.9  71.5  | 2 MJ         |
|  1336    345   3.87   3.90  | -3.02   .16  | .99   .0 1.22  1.0 |  .93 |  .18   .26  | 72.5  75.3  | 6 Rp         |
|  1275    345   3.70   3.77  | -2.06   .10  | 1.33 2.0 1.19  1.2 |  .92 |  .36   .37  | 69.3  69.9  | 7 Su         |
|  1371    345   3.97   3.98  | -4.71   .34  | 1.03  .1 1.39   .8 |  .95 |  .06   .12  | 76.9  77.8  | 8 AT         |
|-----------------------------------------------------------------------------------------------------|
|  1291.9 345.0  3.74   3.81  | -2.54   .14  | 1.03  .2 1.04   .0 |      |  .32        |             | Mean (Count: 8)     |
|    47.7   .0    .14    .10  |   .96   .08  | .15   .9  .21  1.1 |      |  .13        |             | S.D. (Population)   |
|    51.0   .0    .15    .11  |  1.02   .08  | .16  1.0  .22  1.2 |      |  .13        |             | S.D. (Sample)       |
+-----------------------------------------------------------------------------------------------------+
Model, Populn: RMSE .16  Adj (True) S.D. .94  Separation 5.80  Strata 8.07  Reliability (not inter-rater) .97
Model, Sample: RMSE .16  Adj (True) S.D. 1.01  Separation 6.21  Strata 8.62  Reliability (not inter-rater) .97
Model, Fixed (all same) chi-squared: 174.7  d.f.: 7  significance (probability): .00
Model,  Random (normal) chi-squared:  6.5  d.f.: 6  significance (probability): .37
Inter-Rater agreement opportunities: 10836  Exact agreements: 7555 = 69.7%  Expected:  7728.4 =  71.3%
```

Figure 3. Consistency of expert assessment

3.2.4. Validity and reliability test result

The validity and reliability analysis result are presented in the different table. This paper does not present the validity data due to large number of data, while the reliability test result is presented in Figure 4. The validity test data are not presented in this paper. We determined the validity of each item by referring to the criteria to see the outliers is mean square outfit values of 0.5< MNSQ< 1.5. This criterion leaves only 78 valid items from a total of 129 items in the original scale. Although many items were dropped, the remaining valid items were still representative of the aspect and indicators of the instrument. In the aspect of ways of thinking, 23 of 37 items were valid, while in the ways of working, 27 of 39 items were valid.

In the aspect of tools of working, 8 of 21 items were valid, while regarding the last aspect, i.e., living in the world, 20 of 32 items were valid. The range difference between initial and validated scale was 12-14 for each aspect. Although fifty items were dropped, this scale could still be used for the next validation process. As displayed in Figure 3, the reliability score of the developed scale was 0.56 (the reliability with the extreme model). Raters' variability is related to their characteristics, and terms like rater effect, rater error, or rater bias deal with basic raters' variability [35].

```
+------------------------------------------------------------------------------------------------+
| Obsvd   Obsvd   Obsvd  Fair(M)|  +    Model | Infit      Outfit    |Estim.| Correlation |              |
| Score   Count  Average Average|Measure  S.E. | MnSq ZStd  MnSq ZStd |Discrm| PtMea PtExp | Num Aitem    |
|-------------------------------+-------------+----------------------+------+-------------+--------------|
|-------------------------------+-------------+----------------------+------+-------------+--------------|
|  90.5    24.0   3.77   3.83 |  .31   .67 |  .98  .1  1.02  .1 |      |  .24        | Mean (Count: 129) |
|   5.0     .0    .21    .15 | 1.21   .45 |  .44  .7  1.05  .9 |      |  .22        | S.D. (Population)  |
|   5.0     .0    .21    .15 | 1.21   .45 |  .44  .7  1.05  .9 |      |  .23        | S.D. (Sample)      |
+------------------------------------------------------------------------------------------------+
   With extremes, Model, Populn: RMSE .81  Adj (True) S.D. .90  Separation 1.11  Strata 1.82  Reliability .55
   With extremes, Model, Sample: RMSE .81  Adj (True) S.D. .90  Separation 1.12  Strata 1.83  Reliability .56
   Without extremes, Model, Populn: RMSE .56  Adj (True) S.D. .64  Separation 1.13  Strata 1.84  Reliability .56
   Without extremes, Model, Sample: RMSE .56  Adj (True) S.D. .64  Separation 1.14  Strata 1.85  Reliability .56
   With extremes, Model, Fixed (all same) chi-squared:  442.0  d.f.: 128  significance (probability): .00
   With extremes, Model,  Random (normal) chi-squared:   85.0  d.f.: 127  significance (probability): 1.00
```

Figure 4. Reliability of 21st century skills scale

3.3. The 21st century skill scale development results

This research produced a 21st century skill scale for prospective school counselors with a total of 78 valid items. So that it can be used to measure the skill level of the 21st century prospective counselors. The examples result of the instrument written in Table 3.

Table 3. The examples of instrument items

No	Statement item	Answer choices			
		SA	Agree	Disagree	SD
1	I have a partner to talk about my idea				
2	I thought of solutions to the problems with my friends				
3	The difference ideas are an interesting thing to discuss				
4	I think about innovations that accordance with the scientific field				
5	I like to interpret many things				
6	I will analyze everything that happens in my life				
7	I enjoy analyzing things				
8	It was important for me to find out the gist of a series of information				
9	I enjoy to give explanations to others				
10	I have control over myself				

SA=Strongly agree; SD=Strongly disagree

3.4. Discussion

This study aims to develop a 21st-century skill scale for prospective counselors. Counselors play a role in influencing the educational environment and promoting skills that contribute to the students' success in the future [37], [38]. One party who can create a culture of peace educational environment is a counselor. Moreover, various works of literature state that counselors are agents of peace [39]. The presence of 21st-century challenges should be properly balanced by prospective counselors' pivotal role and responsibility. They are responsible for helping students develop their skills and learning experience, making them capable of making career decisions. One of the prospective counselors' roles in students' career success is in the development of self-esteem and realistic attitudes toward future careers [40]. Previous studies have identified the degree to which prospective teachers and counselors affect the development of the 21st-century [41]. It is significantly important for prospective teachers or counselors to develop their own 21st-century skills prior to

imparting these skills to their students. Prospective counselors' skills basically stem from their internals and encompass skills applicable in daily life, such as empathy, harmony, unconditioned regard, and other positive attitudes important for educating and guiding students [42].

In addition to mastering basic skills, prospective counselors are also required to possess more skills relevant to the four aspects of 21st century skills (e.g., way of thinking, way of working, tools for working, and living in the world). The result reflects the importance of the development of 21st-century skill scale for prospective counselors. Items in this study were designed according to the indicators to measure prospective counselors' 21st century skills. Of 129 items, 78 items were found to be valid. Fifty-one items were dropped during the validity test, which may be accounted for by the item difficulty, as presented in the Wright map. The Wright map shows that item difficulty could be estimated and stated in the same linear scale used to state the respondents' performance [33]. The validity test result also relies upon validators' views, which could be reflected in the analysis process [43]. The expert judgment analysis indicates that the eight validators in this study tend to be lenient in assessing each item. This may be accounted for by their inclination to put their rating in certain categories within the scale, and raters' tendency to give higher rates may indicate their leniency [27], [35]. The inter-rater agreement opportunity shows raters' similarity score of 10.836, implying that experts in this study tend to give similar judgments. Validators play a central role in the assessment process, and their variability is manifested in various forms, their mistakes also contribute to the variance of irrelevant construct, which may affect the assessment result [35], [44], [45].

Therefore, further study is required to scrutinize the construct's validity and reliability. A validity test on broader scale is necessary to make this scale representative of Indonesian students. This paper presents a result of a part of a larger, multi years study. The limitation of this study lies in its small scope (i.e., guidance and counseling department students). The study was also limited to expert validation/content validity test. The study implication is twofold. First, the result showed that 21st-century skills are not limited to creative, critical thinking, communication, and collaboration (4Cs). Thus, the 21st-century skill scale could be combined with different perspectives to obtain broader scope. Second, certain aspects in this study were analyzed simultaneously and calibrated into a single linear scale (logit scale). Simultaneous calibration allows the measurement of the seriousness level of a scale with regard to the test takers' capacity, difficulty, and criteria.

4. CONCLUSION

The purpose of this study was to develop a 21st century skills scale for prospective school counselors that had been fulfilled, the instrument had been tested for validity and reliability based on expert judgment. This research produced a 21st century skill scale for prospective school counselors with a total of 78 valid items. So that it can be used to measure the skill level of the 21st century prospective counselors. The scale was developed around four aspects: thinking/learning and innovation skills; ways of work/life skills and careers; job skills/information, media, and technology tools; live in the world. Current research recommends further research on developing 21st century skill-scale instruments for aspiring educators in other scientific fields. It is very important to know that few instruments have been developed to measure the skill level of the 21st century for aspiring educators in various scientific fields. The existence of measurement instruments will help educators know the most reasonable efforts to maximize the skills possessed by students. Further research can use the basis of other theoretical studies to develop the 21st-century skill scale because the skill aspects of the 21st century are quite diverse. This research is limited to prospective school counselors or guidance and counseling students in Indonesia. Developed instruments may not be applicable in other countries and may not apply to prospective counselors outside school settings. The newly involved experts from the Java region have yet to involve experts outside the Java Island. Therefore, there are still many things to be developed and paid further attention to from the limitations of this research.

REFERENCES

[1]   S. M. Kabene, J. Takhar, R. Leduc, and R. Burjaw, "Medical education in the 21st century," *Medical Informatics: Concepts, Methodologies, Tools, and Applications*, vol. 1–4, pp. 178–185, 2008, doi: 10.4018/978-1-60566-050-9.ch015.
[2]   T. Valtonen *et al.*, "How pre-service teachers perceive their 21st-century skills and dispositions: A longitudinal perspective," *Computers in Human Behavior*, vol. 116, no. May 2020, pp. 1–9, 2021, doi: 10.1016/j.chb.2020.106643.

[3]   K. C. Suryandai, Rokhmaniyah, M. Salimi, and S. Fatimah, "Involvement of teachers, parents, and school committees in improving scientific attitudes of elementary school students: Application of Rasch Model Analysis," *International Journal of Educational Methodology*, vol. 8, no. 4, pp. 783–794, 2022, doi: 10.12973/ijem.8.4.783.

[4]   F. Caena and C. Redecker, "Aligning teacher competence frameworks to 21st century challenges: the case for the European digital competence framework for educators (Digcompedu)," *European Journal of Education*, vol. 54, no. 3, pp. 356–369, 2019, doi: 10.1111/ejed.12345.

[5]   Partnership for 21st Century Skills, "Partnership for 21st century skills-Core content integration," Ohio Department of Education, 2015.

[6]   A. J. Rotherham and D. T. Willingham, "'21st century' skills," *American Educator*, vol. 34, pp. 17–20, 2010.

[7]   T. M. Haladyna and M. C. Rodriguez, *Developing and Validating Test Items*. Routledge, 2013, doi: 10.4324/9780203850381.

[8]   H. Boholano, "Smart social networking: 21st century teaching and learning skills," *Research in Pedagogy*, vol. 7, no. 2, pp. 21–29, 2017, doi: 10.17810/2015.45.

[9]   P. Valli, P. Perkkilä, and R. Valli, "Adult pre-service teachers applying 21st century skills in the practice," *Athens Journal of Education*, vol. 1, no. 2, pp. 115–130, 2014, doi: 10.30958/aje.1-2-2.

[10]   A. R. Saavedra and V. D. Opfer, "Teaching and Learning 21st century Skills: Lessons from the Learning Sciences," Paper presented at *the Joint AARE/APERA Conference*, 2012.

[11]   C. Tican and S. Deniz, "Pre-service teachers' opinions about the use of 21st century learner and 21st century teacher skills," *European Journal of Educational Research*, vol. 8, no. 1, pp. 181–197, 2019, doi: 10.12973/eu-jer.8.1.181.

[12]   C. Wilder, "Promoting the role of the school counselor," *Journal of Professional Counseling: Practice, Theory & Research*, vol. 45, no. 2, pp. 60–68, 2018, doi: 10.1080/15566382.2019.1646085.

[13]   C. B. Stone and M. A. Clark, "School counselors and principals: partners in support of academic achievement," *NASSP Bulletin*, vol. 85, no. 624, pp. 46–53, 2001, doi: 10.1177/019263650108562407.

[14]   H. G. McMahon, E. C. M. Mason, and P. O. Paisley, "School counselor educators as educational leaders promoting systemic change," *Professional School Counseling*, vol. 13, no. 2, p. 2156759X0901300, 2009, doi: 10.1177/2156759x0901300207.

[15]   A. P. Maba *et al.*, "Help-seeking intention during the COVID-19 pandemic: sociodemographic factors, anxiety, and fear as predictors," *Minerva Psychiatry*, vol. 64, no. 1, pp. 63–71, 2023, doi: 10.23736/S2724-6612.21.02267-3.

[16]   C. A. Dahir, "Special section school counseling in the 21st century: where lies the future? introduction to the special section," *Journal of Counseling and Development*, vol. 87, no. 1, pp. 3–5, 2009, doi: 10.1002/j.1556-6678.2009.tb00542.x.

[17]   Y. Shtaltovna, "Can a skill be measured or assessed?" *GiLE Journal of Skills Development*, vol. 1, no. 1, pp. 12–24, 2021, doi: 10.52398/gjsd.2021.v1.i1.pp12-24.

[18]   E. Susantini *et al.*, "Developing competency evaluation of pre-service science teachers in industrial revolution 4.0: revealing pedagogic and professional competencies," *International Journal of Educational Methodology*, vol. 8, no. 2, pp. 347–362, 2022, doi: 10.12973/ijem.8.2.347.

[19]   L. Vinet and A. Zhedanov, "A 'missing' family of classical orthogonal polynomials" *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 8. 2011. doi: 10.1088/1751-8113/44/8/085201.

[20]   I. K. Rahman, N. S. Rosidah, and A. M. Tamam, "Development of a scale for measuring the competencies of Islamic counselors," *Islamic Guidance and Counseling Journal*, vol. 6, no. 1, pp. 28–44, 2023, doi: 10.25217/igcj.v6i1.3133.

[21]   R. F. DeVellis, *Scale Development Theory and Applications* (4th Edition). SAGE Publication, 2017.

[22]   D. B. Boyaci and N. Atalay, "A scale development for 21st century skills of primary school students: a validity and reliability study," *International Journal of Instruction*, vol. 9, no. 1, pp. 133–135, 2016, doi: 10.12973/iji.2016.9111a.

[23]   A. Joshi, S. Kale, S. Chandel, and D. Pal, "Likert scale: explored and explained," *British Journal of Applied Science & Technology*, vol. 7, no. 4, pp. 396–403, 2015, doi: 10.9734/bjast/2015/14975.

[24]   J. Mumu, B. Tanujaya, R. Charitas, and I. Prahmana, "Likert scale in social sciences research: problems and difficulties," *FWU Journal of Social Sciences*, vol. 16, no. 4, pp. 89–101, 2022, doi: 10.51709/19951272/Winter2022/7.

[25]   H. Taherdoost, "Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research," *SSRN Electronic Journal*, vol. 5, no. 3, pp. 28–36, 2018, doi: 10.2139/ssrn.3205040.

[26]   R. Heale and A. Twycross, "Validity and reliability in quantitative studies," *Evidence-Based Nursing*, vol. 18, no. 3, pp. 66–67, 2015, doi: 10.1136/eb-2015-102129.

[27]   I. Lamprianou, *Applying the Rasch Model in Social Sciences Using R*. Routledge, 2019. doi: 10.4324/9781315146850.

[28]   K. Lamm, A. Lamm, and D. Edgar, "Scale development and validation: methodology and recommendations," *Journal of International Agricultural and Extension Education*, vol. 27, no. 2, pp. 24–35, 2020, doi: 10.5191/jiaee.2020.27224.

[29]   B. Sumintono, "Rasch model measurements as tools in assessment for learning," *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017)*, vol. 173, 2018, pp. 38–42, doi: 10.2991/icei-17.2018.11.

[30]   J. M. Linacre, *Many-Facets Rasch Measurement*. Institute for Objective Measurement, Inc., 1989.

[31]   K. Kudiya, B. Sumintono, S. Sabana, and A. Sachari, "Batik Artisans' Judgment of Batik Wax Quality and Its Criteria: An Application of the Many-Facets Rasch Model," *Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings*, 2018.

[32]   Maryati, Z. K. Prasetyo, I. Wilujeng, and B. Sumintono, "Measuring teachers' pedagogical content knowledge using many-facet Rasch model," *Cakrawala Pendidikan*, vol. 38, no. 3, pp. 452–464, 2019, doi: 10.21831/cp.v38i3.26598.

[33]   W. J. Boone, "Rasch analysis for instrument development: why, when, and how?" *CBE Life Sciences Education*, vol. 15, no. 4, 2016, doi: 10.1187/cbe.16-04-0148.

[34]   N. Güler, "Analysis of open-ended statistics questions with many facet Rasch model," *Egitim Arastirmalari - Eurasian Journal of Educational Research*, vol. 55, pp. 73–90, 2014.

[35]   T. Eckes, *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments: Second edition*, Berlin, Germany: Peter Lang Verlag, 2011. doi: 10.3726/978-3-653-04844-5.

[36]   D. N. Jackson, "Scale development," in *Encyclopedia of psychology*, Vol. 7, Oxford University Press, 2004, pp. 148–151, doi: 10.1037/10522-060.

[37]   N. Karacan Özdemir and A. Ayaz, "School counseling and guidance programs: examination of career adaptabilities," *Hacettepe Egitim Dergisi*, vol. 35, no. 3. pp. 613–626, 2020, doi: 10.16986/HUJE.2018045437.

[38]   H. M. Alhothali, "Inclusion of 21st century skills in teacher preparation programs in the light of global expertise," *International Journal of Education and Practice*, vol. 9, no. 1, pp. 105–127, 2021, doi: 10.18488/journal.61.2021.91.105.127.

[39]   W. N. E. Saputra, N. Hidayah, M. Ramli, and A. Atmoko, "Development and validation of Indonesian peace of mind scale: the Rasch analysis," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 12, no. 2, pp. 877–885, 2023, doi: 10.11591/ijere.v12i2.24847.

[40] Z. N. Zainudin, S. I. S. S. Abdullah, Y. M. Yusop, and W. N. W. Othman, "STEM education: the career counselor's roles and responsibilities," *International Journal of Academic Research in Business and Social Sciences*, vol. 9, no. 11, pp. 272–283, 2019, doi: 10.6007/ijarbss/v9-i11/6547.

[41] Y. Jia, Y. J. Oh, B. Sibuma, F. LaBanca, and M. Lorentson, "Measuring twenty-first century skills: development and validation of a scale for in-service and pre-service teachers," *Teacher Development*, vol. 20, no. 2, pp. 229–252, 2016, doi: 10.1080/13664530.2016.1143870.

[42] H. Indreswari, Y. Prihatmoko, and N. Nindigraha, "Candidate for personal learning counselors electronic materials on learning environment," *Proceedings of the 1st International Conference on Information Technology and Education (ICITE 2020)*, vol. 508, 2020, pp. 568–572, doi: 10.2991/assehr.k.201214.298.

[43] R. F. Cavanagh and R. F. Waugh, *Applications of Rasch measurement in learning environments research*. SensePublishers Rotterdam, 2011. doi: 10.1007/978-94-6091-493-5.

[44] F. Farrokhi and R. Esfandiari, "A many-facet Rasch model to detect halo effect in three types of raters," *Theory and Practice in Language Studies*, vol. 1, no. 11, pp. 1531–1540, 2011, doi: 10.4304/tpls.1.11.1531-1540.

[45] I. Asakir and D. Hidayati, "Rasch model analysis: teacher commitment indicators by experts judgment," *International Journal of Educational Management and Innovation*, vol. 3, no. 1, pp. 59–73, 2022, doi: 10.12928/ijemi.v3i1.5501.

BIOGRAPHIES OF AUTHORS

Ma'rifatin Indah Kholili is a lecturer in the field of Guidance and Counseling Program at the Faculty of Teacher Training and Education, Sebelas Maret University, Surakarta, Indonesia. Her field of specialization, research areas, publication and presentation cover a wide range of guidance and counselling management program, instrument development, evaluation and supervision, soft skills for prospective counsellor and mobile application for guidance and counselling. She has presented papers at conferences, published articles and papers in various journals. She can be contacted at email: marifatin.ink@staff.uns.ac.id.

Adi Dewantoro is a lecturer in the field of Guidance and Counseling Program at the Faculty of Teacher Training and Education, Sebelas Maret University, Surakarta, Indonesia. His field of interest are behavior change analysis, multicultural counseling, traumatic counseling. He has published articles and papers in various journals. He can be contacted at email: adi_dewantoro@staff.uns.ac.id.

Naharus Surur is a lecturer in the field of Guidance and Counseling Program at the Faculty of Teacher Training and Education, Sebelas Maret University, Surakarta, Indonesia. His field of interest are guidance and counselling management program, evaluation and supervision, guidance and counselling at school. He can be contacted at email: naharus67@staff.uns.ac.id.

Novita Tri Hapsari is a guidance and counseling teacher at junior high school. Now, she is a student at Post Graduate Guidance and Counseling Department, Faculty of Education and Psychology, Yogyakarta State University, Yogyakarta, Indonesia. She is interested in guidance and counseling and mental health research. She also plays a role in the development of applications based on guidance and counseling. She has published article in proceeding. She can be contacted at email novitatrihapsari@student.uns.ac.id.