THE INDIVIDUAL WORK PERFORMANCE SCALE: A PSYCHOMETRIC STUDY AND ITS APPLICATION FOR EMPLOYEE PERFORMANCE

Zulmi Ramdani[1*], Rosleny Marliani[2], Agus Abdul Rahman[3]
[1*, 2, 3]Faculty of Psychology, UIN Sunan Gunung Djati Bandung, Indonesia.
Email: [1*]zulmiramdani@uinsgd.ac.id, [2]roslenymarliani@uinsgd.ac.id, [3]agus.abdulrahman@uinsgd.ac.id

Abstract

Purpose: This research aims to get an individual work performance scale of the modified version which is more acceptable and has a good psychometric property.

Methodology: This study was conducted using a modified measuring instrument approach. Researchers use existing theories and then make items according to the context in which this research will be conducted. The advantage of this modification approach is getting a more comprehensive understanding of the subject we are researching.

Main Findings: Of the 303 employees that were tested in the research, the modification scale of the individual work performance has a good psychometric property with the criteria of reliability more than 0.8 and the fit model that has a good item with RMSEA score (0.062). Other than that, the modification scale of individual work performance shows a good convergent validity with presenteeism and correlate with job burnout.

Applications of this study: The Individual Work Performance Scale has a comprehensive methodology and good psychometric properties. This instrument is appropriate to be the general instrument for seeing the employee situation. In addition, this instrument can also be used as a set of tests that are suitable for use by those who will conduct an assessment of employee performance.

Novelty: This scale was made as a form of development from previous research which did not explore and develop forms of assessment that were specific to an employee's performance. There is not much-related research that focuses on many dimensions that are actually important to assess in employee performance.

Keywords: *Employee Performance, Individual Work Performance, Job Burnout, Performance Appraisal, Presenteeism, Psychometric Properties.*

INTRODUCTION

The development of the use of psychological tests in industrial and organizational fields is one of them influenced by the desire of company leaders to get the best candidates with a good process (Russell & Brannan, 2016). A good process starts by selecting the right measuring instrument and has validity and reliability criteria (Maul, Torres, & Wilson, 2016). These issues were then discussed in the work context until the issuance of a set of principles of validation and standardized use of the test by the Society for Industrial and Organizational Psychology in 1987 (Anastasi & Urbina, 2007). The information conveys the intention that good people have the right to work in a place and a good position or getting the right people (Russell & Brannan, 2016).

Based on an online survey conducted by researchers taking a sample of approximately 64 subjects who worked in the fields of psychology, management, and human resources, as many as 80% of the subjects (composition; 57.8% men and 42.2% women) considered important use of psychological tests in their activities. In addition, other information obtained from the survey are the objectives of using psychological tests such as screening and initial selection of employees, seeing their potential and strengths, knowing weaknesses and problems being experienced by employees, to helping in determining the amount of salary and rewards like what who has the right to receive employees from their jobs. One of the uses of psychological tests is in assessing performance or performance appraisal. Conceptually performance appraisal (PA) is a process or method of assessing employee performance and giving them feedback (Dijk & Schodl, 2015). At present, performance appraisal activities are the most effective method, especially in evaluating the quality and quantity of employee work (Apak, Gümü, Öner, & Gülnihal, 2016), until its contribution to the theoretical and practical developments in the field (Bretz, Milkovich, & Read, 1992).

Performance appraisal is carried out by involving all aspects of the organization and its resources (Berry, 2003). The organization must have a measuring instrument that is in accordance with the conditions and work that is in it (Carlton & Sloman, 1989). In addition, readiness by providing a measuring instrument has an impact on the source of the rater who will provide an evaluation of the indicators of employee performance (Hobson, 1981). So the things that must be considered when going to do this activity include the right measuring instrument, subject (the person being assessed), and the rater or the person who evaluates (Berry, 2003). The challenge to be faced when performing a performance appraisal is to focus on the psychological variables of the performance itself, which can then be seen from the scale or method of assessment used (Hongmei & Tianyi, 2014). This is also revealed from online survey data which shows that the main problem that occurs and makes judgments that subjects do tend to experience bias is the use of methods or psychological

scales that are less precise. Besides that, the assessor's resources also greatly influence the quality of the usual performance measurements (Harari & Rudolph, 2017).

The Individual Work Performance concept presented by Koopmans is the most appropriate instrument to evaluate the performance of an employee. The scale made by Koopmans is compiled using a very complex methodology and goes through various stages of systematic testing both in terms of the theoretical and psychometric strength (Koopmans, Bernaards, & Hildebrandt, 2012). Apart from that, the theoretical basis development on this scale is able to combine all elements of performance into three very important things to measure. These include task performance (main skills), contextual performance (support capabilities), and counterproductive work behavior or things that hinder performance (Koopmans, Bernaards, Hildebrandt, Vet, & Beek, 2014).

An individual work performance measurement tool developed by Koopmans is a set of self-reports that contain statements that are relevant to employment indicators in general (Koopmans et al., 2014). However, there are many challenges that will be faced when using the scale, including the differences in language and culture with the original scale, aspects of performance that may not be measured on a scale, up to the internal factors of the subject that uses it itself (Koopmans et al., 2012). One of the important things to do when going to use the scale is to determine whether the scale is in accordance with the language and culture of the subject that we are going to use (Beaton, Bombardier, Guillemin, & Ferraz, 2000). In addition, the loss factor of the consistency of the psychological scale and the effectiveness factor greatly determines the use of this process (Wild et al., 2005).

The development of individual work performance has been widely used in a variety of work contexts, such as those carried out by Mateen et al., (2017) in analyzing the performance displayed by employees by relating it to health measurements held by individuals. Other results also by evaluating the concepts and psychometric properties that exist on a scale and then doing a theoretical study again regarding the constructs of work performance (Pradhan & Jena, 2017). Most of the studies that use this concept only use existing concepts and then vary in different contexts, so a scale that is in accordance with the place where the measuring instrument is used is needed. In Indonesia there are also a number of studies that try to use the individual concept of work performance, but no one has used and adjusted the measurement according to the appropriate culture and language as well as correlating the measure with other related criteria (Mustafa & Othman, 2010; Mustafa, 2012; Muda, Rafiki, & Harahap, 2014).

Convergent validity refers to a high correlation between measurements in the same construct with different methods (Crocker et al., 2008). The criterion used in this convergent validity is a work performance construct that is more specific to the aspect of presenteeism, while for concurrent validity criteria it uses the Job Burnout variable. Presenteeism is a concept that is a new reference in looking at employee work productivity. This concept refers to the presence of an employee in his work schedule even though he is in a state of illness or health problems (Callen, Lindley, & Victoria, 2013). Presence in unfavorable circumstances is more important than unclear permits or absenteeism(Dew, Keefe, & Small, 2005). Whereas Burnout can be defined as a syndrome or tendency to feel tired physically and emotionally, including the development of negative self-concepts, negative work attitudes, and loss of focus and feelings towards the task and work environment (Pines & Maslach, 1978). The study was conducted to obtain individual work performance scales that have been modified in the Indonesian version and supplemented by testing the convergent and concurrent validity of the scale itself.

LITERATURE REVIEW

Performance: A Psychological Construct

According to Campbell, performance is anything that is done by employees in their work in accordance with the goals and ideals of the organization (Berry, 2003). This can be described as measurable (observable) behavior. Among them are activities - physical activities and activities that may occur in the psychological individual, such as cognitive processes and problem-solving. In addition, Campbell also emphasizes work performance on behavior rather than results, the behavior is in accordance with organizational goals and is multidimensional (Koopmans et al., 2012). Campbell argues that there are three things that occur in performance, namely a declarative knowledge (understanding and knowledge of work tasks), procedural knowledge and skills (abilities and more specific knowledge about the procedures for doing things), and the third is motivation (Berry, 2003).

The measurement of work performance today has brought many changes, especially in creating measuring instruments that have good psychometric properties. According to Koopmans, there are several fundamental reasons why measurement of individual performance is a crucial issue in the scope of assessment of work behavior (Koopmans, Bernaards, Hildebrandt, Cw, & Vet, 2013). First, the globalization of the economy is increasing, which requires integration and independence. Second, individual work performance is needed in an era of economic recession, which is a period in which human activity is greater than the activity of the economy itself. Third, this work performance is very important in determining consistency and maintaining employee work resilience. These three things must be maintained, developed, and optimized when measuring individual work performance (Koopmans et al., 2013). To get comprehensive information about employee performance, another theoretical model that also explains this issue is the performance model presented by Blumberg & Pringle in 1982 (Jewell & Siegall, 1998). This theoretical concept explains that there are at least three important things that

can be the most appropriate indicators if it is to measure how the performance displayed by an employee. These aspects include aspects of capacity, willingness, and opportunity(Blumberg & Pringle, 1982).

Performance Appraisal

One of the important terms to be used is rater. Rater itself is defined as someone who evaluates the performance or performance of others (Berry, 2003). Two things to consider are regarding objective measurement with the subjective measure itself. One of the things when choosing the performance appraisal (PA) method is about the needs of the PA itself along with its advantages and disadvantages. The rater is capable of having the ability and expertise that are appropriate and also they can observe the behavior of the rate performance (the person being assessed). Performance appraisal (PA) is defined as the methods and processes used by the organization to obtain an overview of the level of performance of their employees (DeNisi & Pritchard in Arnăutu & Panc, 2015). This includes measuring performance and feedback on their quality. PA is intended to improve employee performance (Choon & Embi, 2012). According to Nair & Salleh (2015), this goal is achieved through 3 mechanisms: (1) as part of administration in providing reward and punishment as well as salaries, promotions and layoffs; (2) performance feedback through fulfillment strategies, and (3) increasing employee awareness of the facts to be measured.

Changes in issues in the PA begin with the development of psychometric characteristics of the measuring instruments used (Appelbaum, Roy, & Gilliland, 2011). Then regarding the cognitive characteristics of the raters. The more specific the errors of attribution, categorization, stereotypes, and other biases in the process of obtaining information. Effective feedback makes employees concentrate on their performance. Problems that will be faced if employees receive a low rating, they will react negatively, angry, and reject the results of feedback. The most difficulty is that the manager is not enthusiastic about giving negative feedback. It really depends on individual differences in self-esteem, regulatory focus, goal orientation, and performance history. The social context is very influential on the rater, rate, and the relationship between them. It also depends on organizational culture, legal climate, human resource strategy, and organizational goals as well as a leadership style (Dijk & Schodl, 2015).

Current Research

Assessment of an employee's performance is so important in various life contexts. This is because it is directly related to the productivity that will result from their work. Previous research has focused on many factors related to the effectiveness of the use of assessment instruments compared to the psychometric aspects of the measuring instruments they use (Bowden & Sandlund, 2019; Iqbal, Akbar, Budhwar, & Shah, 2019; Rusu, Avasilcăi, & Huţu, 2016). These factors include the existence of an organization's influence which directly affects the assessment process and other activities that are considered to support a person's performance. Although basically in the performance assessment process must consider the related aspects that influence it, most of them forget that the instrument used does not necessarily measure the performance of an employee. The use of good psychometric quality instruments is already widely used, mainly used as the main way to obtain a picture of an employee(Arnăutu & Panc, 2015). Other research is also carried out by looking at other psychological constructs that are directly related to the existence of a leader in the assessment process s(Selvarajan, Singh, & Solansky, 2018), which then indirectly controls the employees' personalities with the various situations they face (Holzman & Valentiner, 2016).

This study was conducted to provide a new alternative to employee performance measurement models that often do not have strong psychometric properties. The instrument which is not strong includes the value of reliability that is not high and validity that is not appropriate so as to produce the wrong measurement. The model of measuring instrument development using a modified approach because it is adapted to the context and phenomena that occur in the field will enable a more comprehensive form of assessment. In addition, the use of good statistical analysis will also help produce a good measurement model. Historically, the individual work performance measuring instrument developed has a good basis so that by making additions and adjustments to the research context it will make it better than the measuring instrument itself.

METHODOLOGY

Participants

Participants involved in this research were campus staff members who were selected using purposive sampling techniques. This technique allows researchers to get subjects that fit the established criteria, namely those who work as administrative staff in the campus environment. Samples involved in this study were 303 subjects, all of whom had already filled out and obtained research approval sheets.

Instruments

Individual Work Performance Scale. The scale to be developed consists of 18 items of the original version that will be modified by the total number of items reaches 40 items. The initial version of the scale has procedures for preparing psychological variables with a good degree of reliability and validity, with details 5 items on aspects of task performance,

8 items on contextual performance aspects, and 5 items on counterproductive aspects of worker behaviour (Koopmans et al., 2012). The measuring instrument itself has received written approval directly by expert via email communication.

Job Burnout (JB) Scale. The scale to be used refers to the modifications made by (Bhratta, 2014) with 3 dimensions tested in 2014 on approximately 87 nurses with alpha 0.922 and item correlations reaching 0.309 to 0.767. Job Burnout Scale JB is 19 items consisting of three dimensions, namely emotional exhausted, cynicism, and inefficacy. The instrument used was approved by the agency where the writing was written by early researchers.

Presenteeism Scale. Measurements are made based on an understanding of the aspects that exist in the employee work environment by using a simple set of questions in the form of a global measurement scale, where respondents determine their work position in the last few weeks whether in the best or worst performance (Kessler et al., 2003). The form of the question is twofold, namely how you assess the performance that is usually done by most workers who have jobs similar to yours and how you assess your work as a whole in the days when you work. For this instrument, researchers made their own measurement tools that were theoretically developed from previous studies.

Research Design

The research procedure begins by modifying the scale of the original version consisting of 18 items (forward translation). This modification consists of the process of translating into Indonesian, then making additions according to the target. The total item target is 40 items, with a response format that has been adjusted to the context in Indonesia. After modification, a trial is carried out to see how the content validity and the different power produced from these items. The next step is to take as many subjects as the criteria to see construct validity using the confirmatory factor analysis method. Then the construct validity is continued by looking at convergent validity with presenteeism variables and looking at concurrent validity with job burnout variables.

RESULTS

Table 1: Reliability Coefficients for Each Variable

Construct	Reliability Coefficients
Task Performance (TP)	0.931
Contextual Performance (CP)	0.914
Counterproductive Work Behavior (CWB)	0.834
Presenteeism	0.827
Job Burnout	0.847

Furthermore, considering the modification approach carried out in this study, the number of items that were previously numbered 18 items were then developed into 40 items arranged based on the suitability and the number of indicators contained in each dimension. The next process when the elements in the initial construction are complete is made in a scaled format that will be tested first by the rater (see table 2).

Table2: Aiken's Validation

Item	V	Item	V	Item	V	Item	V
1	0.86	11	0.94	21	0.81	31	0.77
2	0.77	12	0.89	22	0.90	32	0.87
3	0.74	13	0.72	23	0.78	33	0.82
4	0.85	14	0.84	24	0.89	34	0.72
5	0.84	15	0.86	25	0.79	35	0.76
6	0.86	16	0.79	26	0.80	36	0.70
7	0.88	17	0.88	27	0.93	37	0.80
8	0.89	18	0.76	28	0.88	38	0.81
9	0.81	19	0.84	29	0.79	39	0.78
10	0.83	20	0.78	30	0.88	40	0.81

Note. Referring to the significance of 0.01, then the limit of the value of the validity of the passes is 0.63 (25 raters).

Table 3: Criteria of Fit IWPQ (Individual Work Performance Questionnaire)

Category	TP	CP	CWB	IWPQ	Results
GFI	0.789	0.866	0.920	0.707	Moderate FIT
AGFI	0.724	0.817	0.856	0.672	Moderate FIT

NFI	0.784	0.837	0.876	0.681	Moderate FIT
IFI	0.823	0.883	0.904	0.778	Moderate FIT
PGFI	0.603	0.635	0.511	0.632	Moderate FIT
RMSEA	0.123	0.096	0.112	0.082	Moderate FIT

Some categories such as GFI, AGFI, and IFY are close to the highest limit of 0.9(Darcy & Tracey, 2007). (see figure 1)[Done]
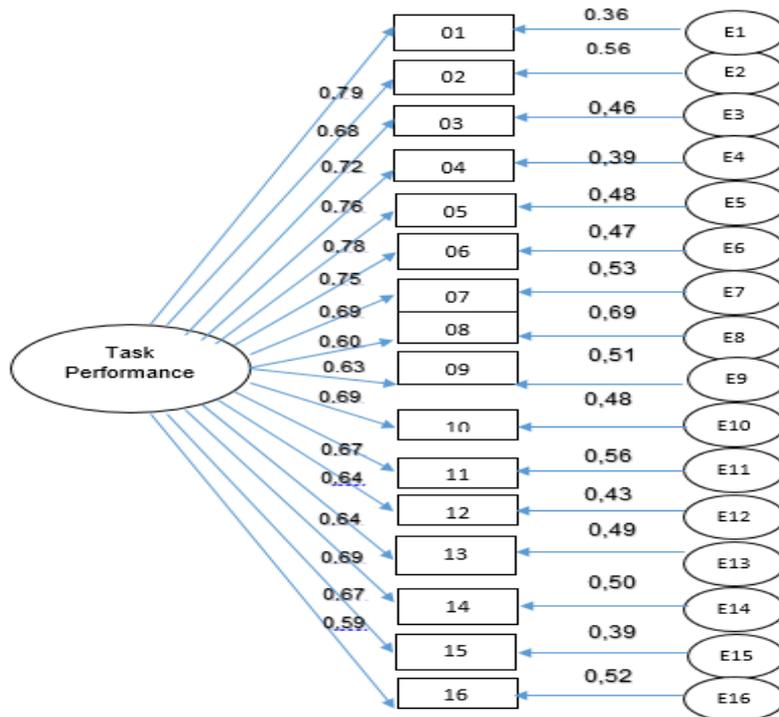

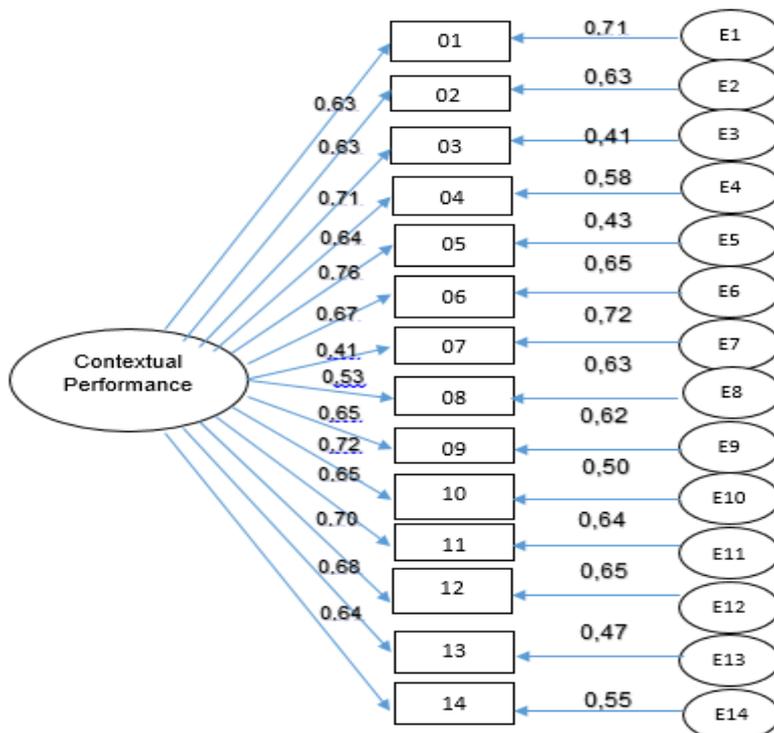
Figure 1: Model Fit Task Performance

Figure 2: Model Fit Contextual Performance

The figure above shows the results of the contextual performance dimension which can be seen that *the loading factors* among items are more than 0.5 (Darcy & Tracey, 2007). In this dimension, the final number of items is 14 items with errors in each item experiencing overloading. Two items that were discarded previously made the model results better. Then below is the result of the model fit dimension counterproductive work behaviour (see figure 3).
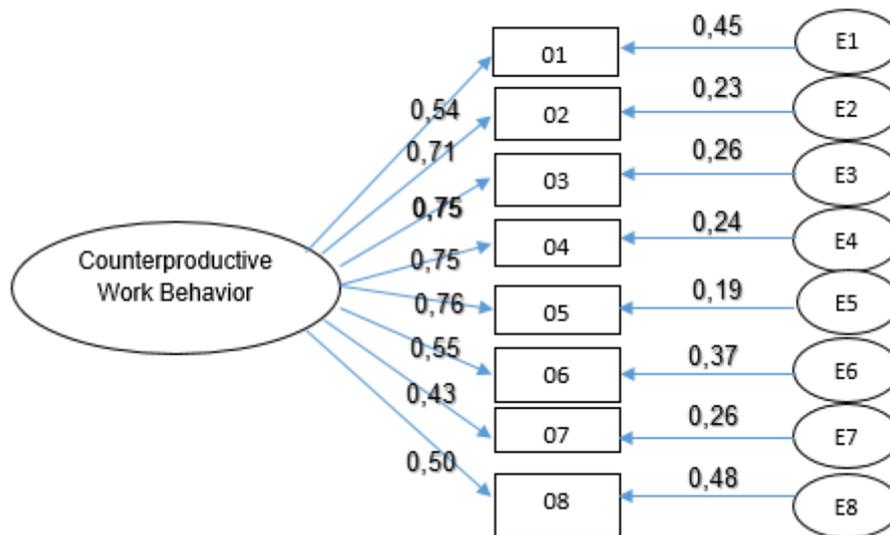

Figure3: Model Fit counterproductive work behavior

The figure above explains the results of the dimension counterproductive work behavior which can be seen that *the loading factor* between items is more than 0.5 (Darcy & Tracey, 2007). In this dimension, the number of fixed items is 8 items with errors in each of the several items experiencing overloading. The models built in this dimension show results that are almost the same as the previous models. After doing factor analysis and testing the fit model, then convergent validity tests were carried out (see table 4).

Table 4: Convergent Validity of IWP with Presenteeism

Dimensions	Absolute Presenteeism	Relative Presenteeism
Task Performance	0.437	0.252
Contextual Performance	0.373	0.234
Counterproductive Work Behavior	-0.311	-0.059

Note. Convergent validity and discriminant validity were obtained from the correlational analysis between scales.

In addition, the validity carried out in this study also tried to see the concurrent validity as an external validity in this study. This validity can be seen from the results of the correlation between IWP and Job Burnout (see table 5).

Table5: Concurrent Validity of IWP with Job Burnout

Dimension	Job Burnout
Task Performance	-0.433
Contextual Performance	-0.342
Counterproductive Work Behavior	0.477

Note. Concurrent validity was obtained by correlating IWP with the Job Burnout scale.

The results above show a high correlation between the three dimensions of IWP with Job Burnout. The difference is that the dimensions of TP and CP have correlation coefficients above 0.3 (respectively -0.433 and -0.334) although the direction is negative. While the counterproductive dimension of work behaviour also produces a high correlation of 0.477 and is positive. The initial two dimensions of IWP have a relationship that is not linear with JB, while CWB has a linear relationship with JB (Job Burnout).

DISCUSSION

The modified approach applied in this study helps researchers to capture a broad picture of the aspects in the field so that they can be involved in making their measuring instruments (Ramdani, 2018). By considering aspects such as psychological characteristics built by constructs and choice of words which are then compared to one another, then with this translation result, it can be concluded that the first translator is more accepted than the translation compared to the second translator. The simple reason is that the experience and quality of the first translator are more numerous, and its superiority is being able to present sentences well and be well understood. But the collaboration between the two and systematic synthetics makes the translation results in better (Bullinger et al., 1998).

The modification process is to increase the number of items on each indicator so that it is more proportional and also changes the format of the response into a format that is more acceptable to the respondent. Each indicator in each dimension will be added to the same amount so that each indicator consists of four items. Based on the results of content analysis on 40 modified items, the validation criteria were 0.63. This limit becomes the standard for passing items that will be tested in the trial. Of the 40 items made in the modification phase, all items have validation values above 0.63. High item-content relevance indicators become an important foundation in building a strong construct in measurement because there is a good connection between the fellow item (Dunn & Bouffard, 2009). The right appraiser has the same understanding and perspective regarding items relevant to the construct. Unlike those who choose the wrong rater, it causes cognitive failure (Wallace & Chen, 2005).

The results of discrimination power tests on 40 respondents produced a very high value of different power in each item. The data obtained shows that each dimension has a value of average discrimination power which varies greatly from task performance (rix = 0.542), contextual performance (rix = 0.465), and counterproductive work behavior (rix = 0.705). This explains that the items in these three dimensions are able to distinguish between high-performance individuals and low-performance individuals. The three reliability coefficients generated from the three dimensions show almost a very high coefficient because it reaches more than 0.7. Similarly, the results shown in other variables such as job burnout reached 0.847 and presenteeism variables reached 0.827. Some reasons that cause high or low reliability are measurement errors, number of items, number of subjects, score homogeneity, and the correlation between items in the test (Anastasi & Urbina, 2007; see also Vet, Mokkink, Mosmuller, & Terwee, 2017).

Confirmatory factor analysis is used when the researcher already has a clear conceptual picture of the construct being measured (Thompson, 2004). The number of respondents used in the study amounted to 232 subjects. This amount has met the requirements for factor analysis, which is a sample of more than 100 (Gudono, 2016). Based on these criteria, the factor analysis study in the study resulted in a magnitude of 0.906 which indicates a very good result so that it can be said that the exact factor analysis is used. The value of the KMO obtained in the study is one of them due to the large number of samples that can represent each of the factors determined (Meng & Jin, 2017). In addition, loading factors in these items produce a significant relationship so that it is appropriate to conduct an in-depth analysis of each of the factors (Gudono, 2016).

There are several important things that happen by considering the fit model in the SEM (Structural Equation Modeling) results. First, the bigger or closer to the criteria specified for each category, then this will also affect item validity and overall reliability (Turner & Allen, 2018). The second classification of the goodness of fit index criteria is based on the output that produces criteria based on absolute fit, namely scores on GFI and RMSEA, then incremental fit scores on AGFI, NFI, CFI, IFI, and RFI, while the last is criteria in parsimonious fit, namely PGFI, PNFI, and AIC. Referring to the absolute index category, the RMSEA IWP results reached 0.062 and GFI around 0.707 so that the IWP model is a model that is fit and shows good sharing items.

The results of the previous exposure show that the correlation calculations for each dimension prove that there is convergent validity between the IWP and presenteeism. The correlation coefficients reached 0.437 (TP), 0.373 (CP), and -0.311 (CWB) on absolute presenteeism, and the results were 0.252 (TP), 0.234 (CP), and -0.059 (CWB) in the relative presenteeism dimension. Such results show a high correlation, where theoretically task performance and contextual performance will be of higher value than counterproductive work behaviour. Similarly, the direction of the correlation shown by the first two dimensions is positive, and for the negative direction dimension. Higher results indicated by absolute scores show that the performance construct does measure something that is actual, rather than trying to compare themselves with the performance shown by others as done in relative presenteeism (Koopmans et al., 2014; see also Wang et al., 2003). Other convergent studies are also supported by findings from Rehan at all (2017), which shows a negative correlation between aspects of CWB and presenteeism (Rehan, Iqbal, Fatima, & Nawab, 2017). Another explanation that indeed supports the results of the study on this convergent test is that the correlation coefficient is not too high, even though there is a significant relationship. Correlation coefficients that do not reach 0.7 and above are caused by the measured construct is very individual, so that when the subject gives a lot of bias assessment that occurs in it.

The relationship between task performance and job burnout resulted in a high but negative trending correlation (-0.433). Likewise with the results of the contextual performance correlation with job burnout which is (-0,342). However, the counterproductive work behavior correlation with burnout jobs is positive, which is 0.477. These results are theoretically in accordance with the construct that was in the previous discussion. The two constructs are two constructs built in the context of work, but the directions of both are not linear. This indicates that when an individual person experiences job

burnout conditions their performance decreases and vice versa if their work performance is high then they are not experiencing job burnout.

The overall results of the study produce an Individual Work Performance scale that can be used as a tool in determining the conditions of employee performance. First, this instrument is based on theoretical constructs with good psychometric analysis, so that the level of validity and the results obtained are more maximal. This will make the measurement results more leverage because the use of measuring instruments is well standardized. Second, related to the performance appraisal process, this measuring instrument is able to see the psychological condition of employees, which includes aspects that cannot be seen directly, such as the basic abilities of employees, their relationships with each other, and how they perceive themselves in their organization. Everything that cannot be directly observed by the assessor can be obtained from the results of the analysis of this measuring instrument (Koopmans, 2015). This means that by using this tool as a performance measurement tool, characters such as awareness and mental health of employees can be identified from the start. This will help open up how the actual conditions happened to employees in recent times, which will then impact the need for counseling and the use of methods to improve their performance. From this result, a new, more practical framework will be made, including the use of guidelines that have been adjusted to the working conditions of the employees.

Third, as a human assessment instrument that requires data integration, the scale of the IWP cannot necessarily be used as the main tool. This indicates that this gauge will be more effective and produce more valid measurements if used as an instrument that is integrated with other instruments (Anastasi & Urbina, 2007), such as the KPI (Key Performance Indicator) method, conducting raters based assessments and using other qualitative methods such as interviews. Because this measuring instrument is a measuring instrument that cannot stand alone as a research instrument so that another set of tests is needed, which is also valid and has the same purpose in measuring employee performance. So that the right collaboration between measuring devices will result in a more comprehensive conclusion of the performance.

CONCLUSION

The modified version of the individual work performance (IWP) scale has good psychometric strength seen from the high-reliability coefficient, the fit model of the construct it builds, and the validity criteria both in terms of the construct and the external. The reliability coefficient for each dimension, task performance is 0.931, contextual performance is 0.904, and counterproductive work behaviour is 0.834. The fit model built-in IWP has good sharing of items so that it shows the right criteria to be used as a general self-report. Besides that, IWP also correlates convergently with presenteeism and has concurrent validation with the scale of job burnout. Therefore, the modified version of IWP is highly recommended for use in assessing and evaluating employee work performance in general, especially those in the academic environment.

LIMITATION

Some weaknesses that occur during the research process contribute to the non-maximization of results, including the absence of subject assistance when the administrative process is scaled so that the bias and possible error in the assessment are greater. The tendency to avoid extreme responses to sensitive parts of the scale also occurs, as well as the period of work that is seen as long enough to remember all the processes and activities that have been carried out in work over the past three months. The use of more modern methods and also more and varied samples is also a consideration in the analysis.

PRACTICAL APPLICATION

This study produces a measuring tool that is practically biased to be used by all agencies that aim to conduct performance evaluations of an employee's performance. With standardized measuring instruments, it is possible that this measuring instrument can be used in various contexts related to individual performance. To enrich the assessment of this measuring instrument, the researcher suggests combining this measurement with any other test that does measure other things that are not psychologically measurable.

ACKNOWLEDGMENT

REFERENCES

1. Anastasi, A., & Urbina, S. (2007). *Tes Psikologi : Psychological Testing (Terjemahan)* (Ketujuh). Jakarta: PT Indeks.
2. Apak, S., Gümü, S., Öner, G., & Gülnihal, H. (2016). Performance appraisal and a field study. *Procedia - Social and Behavioral Sciences*, *229*, 104–114. doi.org/10.1016/j.sbspro.2016.07.119
3. Appelbaum, S. H., Roy, M., & Gilliland, T. (2011). Globalization of performance appraisals : theory and applications. *Management Decision*, *49*(4), 570–585. doi.org/10.1108/00251741111126495
4. Arnăutu, E., & Panc, I. (2015). Evaluation criteria for performance appraisal of faculty members. *Procedia - Social and Behavioral Sciences*, *203*, 386–392. doi.org/10.1016/j.sbspro.2015.08.313

5. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures, *25*(24), 3186–3191.

6. Berry, L. M. (2003). *Employee selection*. Canada: Thomson & Wadsworth.

7. Bhratta, W. B. (2014). *Pengaruh Pelatihan Hatha Yoga Terhadap Job Burnout Masa Kerja Sebagai Kovariabel*. Yogyakarta.

8. Blumberg, M., & Pringle, C. D. (1982). The missing opportunity in organizational research: Some implications for a theory of work performance. *The Academy of Management Review*, *7*(4), 560–569.

9. Bowden, H. M., & Sandlund, E. (2019). Knowledge talk in performance appraisal interviews. *Learning, Culture and Social Interaction*, *21*(December 2018), 278–292. doi.org/10.1016/j.lcsi.2019.03.012

10. Bretz, R., Milkovich, G., & Read, W. (1992). The current state of performance appraisal research and practice: concerns, directions, and implications. *Journal of Management*, *18*(2), 321–352.

11. Bullinger, M., Alonso, J., Apolone, G., Leplège, A., & Sullivan, M. (1998). Translating health status questionnaires and evaluating their quality : The IQOLA project approach. *Journal of Clinical Epidemiology*, *51*(11), 913–923.

12. Callen, B. L., Lindley, L. C., & Victoria, P. (2013). Health risk factors associated with presenteeism in the workplace. *JOEM*, *55*(11), 1312–1317.doi.org/10.1097/JOM.0b013e3182a200f4

13. Carlton, I., & Sloman, M. (1989). Performance appraisal in practice. *Human Resource Management Journal*, *2*(3), 80–94.

14. Choon, L. K., & Embi, M. A. (2012). Subjectivity, organizational justice and performance appraisal: Understanding the concept of subjectivity in leading towards employees' perception of fairness in the performance appraisal. *Procedia - Social and Behavioral Sciences*, *62*, 189–193. doi.org/10.1016/j.sbspro.2012.09.030

15. Crocker, L., Crocker, L., Alglna, J., Staudt, M., Mercurio, S., Hintz, K., & Walker, R. A. (2008). *Introduction to classical and modern test theory*. United State of America: Cengage Learning.

16. Darcy, M. U. A., & Tracey, T. J. G. (2007). Circumplex structure of holland ' s RIASEC interests across gender and time. *Journal of Counseling Psychology*, *54*(1), 17–31. doi.org/10.1037/0022-0167.54.1.17

17. Dew, K., Keefe, V., & Small, K. (2005). Choosing to work when sick : Workplace presenteeism. *Social Science & Medicine*, *60*, 2273–2282. doi.org/10.1016/j.socscimed.2004.10.022

18. Dijk, D. Van, & Schodl, M. M. (2015). Performance appraisal and evaluation. *International Encyclopedia of Social & Behavioral Sciences*, *17*, 716–721. doi.org/10.1016/B978-0-08-097086-8.22034-5

19. Dunn, J. G. H., & Bouffard, M. (2009). Assessing item in sport psychology research : Issues and recommendations. *Measurement in Physical Education and Exercise Science*, *3*(1), 37–41. doi.org/10.1207/s15327841mpee0301

20. Gudono. (2016). *Analisis data multivariat* (Keempat). Yogyakarta: BFFE-Yogyakarta.

21. Harari, M. B., & Rudolph, C. W. (2017). The effect of rater accountability on performance ratings : A meta-analytic review. *Human Resource Management Review*, *27*(1), 121–133. doi.org/10.1016/j.hrmr.2016.09.007

22. Hobson, C. J. (1981). Clarifying performance appraisal criteria. *Organizational Behavior and Human Performance*, *188*, 164–188.

23. Holzman, J. B., & Valentiner, D. P. (2016). Self-focused attention affects subsequent processing of positive (but not negative) performance appraisals. *Journal of Behavior Therapy and Experimental Psychiatry*, *50*, 295–302. doi.org/10.1016/j.jbtep.2015.10.008

24. Hongmei, Z., & Tianyi, Z. (2014). The study of the differences of appraisal results by self and appraisal results by Line-manager in performance appraisal. *Journal of Statistics and Management*, (December 2014), 37–41. doi.org/10.1080/09720510.2010.10701497

25. Iqbal, M. Z., Akbar, S., Budhwar, P., & Shah, S. Z. A. (2019). Effectiveness of performance appraisal: Evidence on the utilization criteria. *Journal of Business Research*, *101*(May 2018), 285–299. doi.org/10.1016/j.jbusres.2019.04.035

26. Jewell, L., & Siegall, M. (1998). *Psikologi industri/organisasi modern*. (A. Danuyasa, Ed.) (Second edi). USA: Arcan.

27. Kessler, R. C., Barber, C., Beck, A., Berglund, P., Ba, M., Cleary, P. D., … Ustun, T. B. (2003). The world health organization health and work performance questionnaire (HPQ). *JOEM*, *45*(2), 156–174. doi.org/10.1097/01.jom.0000052967.43131.51

28. Koopmans, L. (2015). *Individual work performance questionnaire : Instruction manual*. *TNO innovation for Life - VU University Medical Center*.

29. Koopmans, L., Bernaards, C., & Hildebrandt, V. (2012). Development of an individual work performance questionnaire. *International Journal of Productivity and Performance Management*, *62*(No. 1, 2013), 6–28. doi.org/10.1108/17410401311285273

30. Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Cw, H., & Vet, D. (2013). Measuring individual work performance : Identifying and selecting indicators. *Galley Prof*, *00*, 1–10. doi.org/10.3233/WOR-131659

31. Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Vet, H. C. W. De, & Beek, A. J. Van Der. (2014). Construct validity of the individual work performance questionnaire. *JOEM*, *56*(3), 331–337.

doi.org/10.1097/JOM.0000000000000113

32. Mateen, B. A., Doogan, C., Hayward, K., Hourihan, S., Hurford, J., & Playford, E. D. (2017). Systematic Review of Health-Related Work Outcome Measures and Quality Criteria-Based Evaluations of Their Psychometric Properties. *Archives of Physical Medicine and Rehabilitation*, *98*(3), 534–560.doi.org/10.1016/j.apmr.2016.06.013

33. Maul, A., Torres, D., & Wilson, M. (2016). On the philosophical foundations of psychological measurement. *Measurement*, *79*, 311–320.doi.org/10.1016/j.measurement.2015.11.001

34. Meng, L., & Jin, Y. (2017). A confirmatory factor analysis of the utrecht work engagement scale for students in a chinese sample. *Nurse Education Today*, *49*, 129–134. doi.org/10.1016/j.nedt.2016.11.017

35. Muda, I., Rafiki, A., & Harahap, M. R. (2014). Factors Influencing Employees ' Performance : A Study on the Islamic Banks in Islamic Science University of Malaysia University of North Sumatera. *International Journal of Business and Social Sience*, *5*(2), 73–81.

36. Mustafa, M. N. (2012). Teacher perspectives on work performance: A review in high school Riau province, Pekanbaru, Indonesia. *Research Journal of Applied Sciences*. doi.org/10.3923/rjasci.2012.458.465

37. Mustafa, M. N., & Othman, N. (2010). The Effect of Work Motivation on Teacher ' s Work Performance in Pekanbaru Senior High Schools , Riau Province , Indonesia. *Sosiohumanika*, *3*(2), 259–272.

38. Nair, M. S., & Salleh, R. (2015). Linking performance appraisal justice, trust, and employee engagement: A conceptual framework. *Procedia - Social and Behavioral Sciences*, *211*, 1155–1162. doi.org/10.1016/j.sbspro.2015.11.154

39. Pines, A., & Maslach, C. (1978). Characteristics of staff burnout in mental health settings. *Hospital & Community Psychiatry*, *29*(4), 233–237.

40. Pradhan, R. K., & Jena, L. K. (2017). Employee performance at workplace: conceptual model and empirical validation.*Business Perspectives and Research*, *5*(1), 69–85. doi.org/10.1177/2278533716671630

41. Ramdani, Zulmi. (2018). Construction of academic integrity scale. *International Journal of Research Studies in Psychology*, *7*(1), 87–97. doi.org/10.5861/ijrsp.2018.3003

42. Rehan, M., Iqbal, M. Z., Fatima, A., & Nawab, S. (2017). Organizational cynicism and its relationship with employee's performance in teaching hospitals of pakistan. *International Journal of Economics & Management Sciences*, *6*(3). doi.org/10.4172/2162-6359.1000413

43. Russell, S., & Brannan, M. J. (2016). Getting the right people on the bus: Recruitment , selection and integration for the branded organization. *European Management Journal*, *XXX*, 1–11. doi.org/10.1016/j.emj.2016.01.001

44. Rusu, G., Avasilcăi, S., & Huţu, C.-A. (2016). Organizational context factors influencing employee performance appraisal: A research framework. *Procedia - Social and Behavioral Sciences*, *221*, 57–65. doi.org/10.1016/j.sbspro.2016.05.090

45. Selvarajan, T. T., Singh, B., & Solansky, S. (2018). Performance appraisal fairness, leader member exchange and motivation to improve performance: A study of US and Mexican employees. *Journal of Business Research*, *85*(April 2016), 142–154. doi.org/10.1016/j.jbusres.2017.11.043

46. Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington: American Psychological Association.

47. Turner, M. J., & Allen, M. S. (2018). Confirmatory factor analysis of the irrational performance beliefs inventory (iPBI) in a sample of amateur and semi-professional athletes. *Psychology of Sport & Exercise*. doi.org/10.1016/j.psychsport.2017.11.017

48. Vet, H. C. W. De, Mokkink, L. B., Mosmuller, D. G., & Terwee, C. B. (2017). Spearman-brown prophecy formula and cronbach's alpha: Different faces of reliability and opportunities for new applications. *Journal of Clinical Epidemiology*. doi.org/10.1016/j.jclinepi.2017.01.013

49. Wallace, J. C., & Chen, G. (2005). Development and validation of a work-specific measure of cognitive failure : Implications for occupational safety. *Journal of Occupational and Organizational Psychology*, *78*, 615–632. doi.org/10.1348/096317905X37442

50. Wang, P. S., Beck, A., Berglund, P., Leutzinger, J. A., Pronk, N., Schenk, T. W., … Kessler, R. C. (2003). Chronic medical conditions and work performance in the health and work performance questionnaire calibration surveys. *JOEM*, *45*(12), 1303–1311. doi.org/10.1097/01.jom.0000100200.90573.df

51. Wild, D., Grove, A., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health*, *8*(2), 94–104. doi.org/10.1111/j.1524-4733.2005.04054.x