Measurement Scales of Administrative Services and Academic Culture in Universities Based on Item Response Theory

Burhanuddin(✉), Bambang Budi Wiyono, Maisyaroh, Achmad Supriyanto, and Nur Luthfi Ardhian

Universitas Negeri Malang, Malang 65145, East Java, Indonesia
burhanuddin.fip@um.ac.id

Abstract. Examine the performance of the universities in providing administrative services to customers and building academic culture need to be measured using appropriate measures scales. Unidimensional factor is an important criterion that must be deserved in scale development for this purpose. This paper reports a study focusing on measuring university performance on the two elements referring item response theory using the Rasch Model. Results show that the validated items of "university administration services" and "student academic culture" can be classified as acceptable instrument items. They are then claimed as having predictability performance in measuring the researched variables.

Keywords: Administrative Services · Academic Culture · Unidimensional · Measurement · Scale

1 Introduction

Establishing universities with international reputation needs accountable customer services and supportive academic culture (Ahmad, 2015; Burhanuddin, 2017; Jones, Lefoe, Harvey, & Ryland, 2012). The two elements are parts of the universities' performance variables that need to be assessed regularly to find out how university organizations deliver services and create academic cultures for their students (Basir, Davies, Douglas, & Douglas, 2017; Smith, Smith, & Clarke, 2007).

In order to examine the performance of the universities in carrying out the two task components, reliable measures of both variables are required. This stage can be obtained through developing and validating the scales involving appropriate statistical procedures (Hair_Jr., Black, Babin, & Rolph E, 2016; Levin, Rastogi, Rubin, & Siddiqui, 2014). This paper highlights related study findings how the constructs of administrative services and academic culture are derived and validated in referring to item response theory (IRT) (Wu & Adams, 2007).

This research involves a large number of variables representing administrative services and academic culture of the universities in Indonesia. To facilitate the measurement development, it deems for the authors' team to reduce variables' number of the researched

B. B. Wiyono et al. (Eds.): ICEMT 2022, ASSEHR 727, pp. 583–592, 2023.
https://doi.org/10.2991/978-2-494069-95-4_67

components into a relatively small number of the component groups. However, results of this simplification step still retain basic information possessed by the original main variables. The most appropriate statistical tool to solve such issues is Principal Component Analysis (PCA). This statistical tool is one of the mathematical procedures that seeks to transform a number of variables which are estimated to be related to each other (correlated variables) into a small number of uncorrelated variables, or called as "principal components" (Burhanuddin & Sunarni, 2016; Hair_Jr., et al., 2016).

Unidimensional measurement is an important criterion that must be deserved by measurement tools, especially scales. The reason is that a concept being measured has its unidimensional characteristic. On the other hand, if in reality the concept being measured has multidimensional characteristics, then the measuring instrument must also be multidimensional. When we measure the height and weight of a person, for example, it is certain that the measuring instrument needed is in the form of a unidimensional measuring scale because only a definite measuring instrument is needed to determine the height and weight. If the achievement is combined as a single achievement, then it can be called unidimensional, so that only a single-dimensional measuring instrument is needed for example the measurement scale of student achievement "very high, high, sufficient, low". However, if the student's ability is seen from his achievement that has two dimensions for example of verbal and mathematical skills, the measuring instrument (scale) needed is more precisely as multidimensional.

Items of the instrument development described in this paper are treated as elements or indicators that function to support the formation of a 'unidimensional' measurement scale. The reason is that all items designed in the measurement scales used are expected to measure specifically concepts, constructs or latent variables that are unidimensional. This means that these latent constructs or variables are only designed to measure specific behaviours or concepts that can only be measured by each specific item. Thus, a concept that has been measured and explained by items (indicators) will not be measured by other groups of variables. Wu and Adam (2007: 21) stated "there is one latent variable of interest, and the level of this latent variable is the focus of the measurement". So, it is imperative to meet the unidimensional characteristics.

To ensure the items are unidimensional, the study reported in this paper requires item analysis using the Rasch Model approaches to several variables, especially aspects of university staff services and student academic culture within the university environment.

Unidimensional nature of measurement items is tested through experiment process carried out with the help of the Rach model technique using Conquest 2.0 software (Wu, Adams, & Haldane, 2005). This technique is expected to be able to estimate the values of fit statistics and thresholds for each item. Results of the calculation of these values are very useful for determining item quality level. This technique statistically aimed at identifying items that have capacities to measure accurately each construct that represents its items. Fit statistics values produced in this study include estimates, MNSQR (mean square), t values, and thresholds (γ) as presented in the tables displaying results of Rasch model. Thresholds (γ) is defined as a cutting point that divides a continuum of respondent behaviours in answering each question item into a range of score regions (Wu & Adams, 2007). This technique can clearly and consistently divide the distance of each possible answer option or perception on a question. Example of this distance is stated with a score

Table 1. Interpretation of Mean Square (MNSQ)

Mean-square values	Item interpretation for the measurement model
>2.0	Item distorts the measurement system that is probably caused by one or two observations in the survey.
1.5–2.0	Unproductive item, but not degrading the measurement model.
0.5–1.5	Productive item for the measurement model
< 0.5	The item is less productive but not degrading for the measurement model. This may produce misleading high reliability and separation coefficients.

Table 2. Interpretation of t values

Standardised values	Item interpretation for the measurement model
≥3	Data are unexpected if they fit the model (perfectly), so they probably do not. However, with large sample size, the substantive misfit may be small and its effect on the model is not significant.
2.0–2.9	Data noticeably unpredictable
−1.9–1.9	Data have reasonable predictability.
≤2	Data are too predictable. Other "dimensions" may constrain the response patterns.

of 5 = Strongly Agree, score 4 = Agree, and so on according to the width of the Likert scale distance that is consistently used in research.

$$\frac{\sum_{n} z_{ni}^{2}}{N} = \frac{1}{N}\sum_{n}\frac{\left(x_{ni} - E\left(x_{ni}\right)\right)^{2}}{Var\left(x_{ni}\right)}$$

The fit t statistic can be interpreted as a normal deviation with a calculated average of "0" and a standard deviation of "1". This value is a transformation of men-square by taking into account arithmetic mean and variance of statistical mean-square fit. In a simple computation process that the fit value is solved by the following formulas (Wu & Adams, 2007).

In the view of item response theory (Item Response Theory) the MNSQR value closes to number "1" indicating that the item can be categorized according to the proposed model (fits the model well) or the observation results are in accordance with what is expected by the measurement model (Table 2).

On the other hand, those that are far above and below the range of "1" show that items have a relatively poor level of differences to the model (poor fit to the model). If these happen, then the items can be assessed as problematic or categorized as miss-fit or

otherwise over-fit items. Guidelines for interpreting statistical values obtained through the Rasch model analysis technique follow the guidelines proposed by Linacre (Linacre, 2009) as presented in Table 1.

2 Methods

This study uses a development research approach, which systematically involves multidimensional theoretical and empirical studies to produce intangible products in the form of a measurement model to examine the performance of universities' administrative services and student academic culture (Kelly, Baek, Lesh, & Bannan-Ritland, 2008; Richey, Klein, & Nelson, 2004). The theoretical model of this research was developed so that it can be used as a guideline for researching and describing complex relationships between the variables studied (Cramer, 2003; Hair_Jr. et al., 2016). At the same time, it is intended to close the gap so far in obtaining information on how the measures of administrative services and university academic culture can be developed and validated properly following the Rash model.

A total of 350 participants have been included as research samples taken from 5 (five) universities in Indonesia. The target universities were chosen based on the considerations of the representativeness of the Java and outside Java regions, the distance to the location, and the willingness of sample members to be the target of data collection. In addition, it is estimated that each has a unique organizational culture that can affect the performance of staff leadership in the university organizational environment (Selim Jahan & Eva Jespersen, 2015; UNESCO, 2014; Varghese et al., 2014). The size of the number was determined purposively based on the sample size table (Creswell, 2005) by considering the level of confidence in the sampling variability. The tolerable sampling error rate is 5%, which means that only 5% of the mean obtained from the sample (sample mean) will differ from the true population mean(Creswell, 2005).

The data collection technique used two sets of survey questionnaires, namely a set for lecturers and employees' respondents. Each item is designed using a Likert scale format with five alternative choices (five-point Likert scale). Instrument validation, especially items on the measurement scale was carried out to test whether the items in each measurement scale developed get valid answers. Content validation was done where all items were consulted with the intention stated in the research proposal or theoretical model (Creswell, 2014). This strategy is intended to consider the extent of its suitability with the research objectives. Followed by applying construct validation process using the Rasch model technique. The goal is to identify whether the items actually measure exactly what each construct represents in the instrument's theoretical model (Creswell, 2014). The computation of item statistical values was carried out with the help of the ConQuest 2.0 software program introduced by Wu et al. (2007). Data were then analysed using both descriptive and inferential analysis techniques. The descriptive stage was intended to describe general state of data where mean, variance, standard deviation values are calculated (Gray, 2009; Stevens, 2009). This was followed by inferential analysis stage to examine relations or effects among variables and draw conclusions based on sample responses.

3 Result and Discussion

Implementation of Principal Component Analysis (PCA) in the measurement model grouped many items into sub-variables with a smaller number of items. This procedure successfully transformed correlated variables into a small number of uncorrelated ones. Variables and items that were successfully validated are presented in Table 3. These items are grouped into each latent variable based on correlation values or item loadings. In other words, they are items that correlate with their representative latent variables or contribute significantly to the variance produced by each item on these latent variables. Thus, they form each sub-variable with a number of items that function as indicators or observed variables.

Results of the analysis of these items are presented in Table 3. A total of 28 measurement scale items for "university administration services" have been successfully developed based on the results of this study. The items for this scale are then distributed into 3 (three) sub-scales including: (1) employee attitudes in administrative services, (2) direct communication behaviour with students, and (3) communication behaviour when receiving calls or making calls. For the variable items "student academic culture" there are 12 items prepared (consisting of 4 items of time use culture and 8 items of academic

Table 3. Parameter estimates of item response model for administrative services and student academic culture

			Weighted Fit						
Variable	Estimates	Error	MNSQR	CI	t	Item thresholds			
Staff attitude In providing services									
Sserve116	0.085	0.075	0.95	0.80, 1.20	-0.5	-2.91	-0.79	0.85	3.18
Sserve117	0.127	0.075	0.92	0.80, 1.20	-0.7	-2.86	-0.75	0.90	3.22
Sserve118	0.259	0.075	1.21	0.80, 1.20	1.9	-2.73	-0.62	1.03	3.35
Sserve119	-0.033	0.075	0.98	0.80, 1.20	-0.2	-3.02	-0.91	0.73	3.06
Sserve120	0.088	0.075	1.37	0.80, 1.20	3.2	-2.90	-0.79	0.86	3.18
Sserve121	0.132	0.075	0.80	0.80, 1.20	-2.1	-2.86	-0.74	0.90	3.23
Sserve123	0.323	0.075	1.03	0.80, 1.20	0.3	-2.66	-0.55	1.09	3.41
Sserve124	-0.303	0.075	0.82	0.80, 1.20	-1.8	-3.29	-1.18	0.47	2.79
Sserve129	-0.288	0.075	1.15	0.80, 1.20	1.4	-3.27	-1.16	0.48	2.80
Sserve131	-0.390*	0.226	0.92	0.80, 1.20	-0.7	-3.38	-1.27	0.38	2.70
Direct communication									
Sserve122	-0.149	0.080	1.08	0.80, 1.20	0.8	-3.65	-1.01	0.78	3.28
Sserve125	0.078	0.080	0.95	0.80, 1.20	-0.4	-3.42	-0.78	1.01	3.51
Sserve126	-0.225	0.080	1.13	0.80, 1.20	1.2	-3.72	-1.08	0.70	3.20
Sserve127	0.259	0.079	1.03	0.80, 1.20	0.4	-3.23	-0.59	1.19	3.69
Sserve128	0.134	0.079	0.71	0.80, 1.20	-3.0	-3.37	-0.72	1.06	3.56
Sserve130	0.046	0.079	0.82	0.80, 1.20	-1.8	-3.45	-0.81	0.98	3.47
Sserve132	-0.204	0.079	1.12	0.80, 1.20	1.1	-3.70	-1.06	0.73	3.22
Sserve133	0.060*	0.210	0.97	0.80, 1.20	-0.2	-3.44	-0.80	0.98	3.48
Telephone communication									
Sserve134	0.514	0.082	1.12	0.79, 1.21	1.2	-2.95	-0.60	1.42	4.15
Sserve135	0.210	0.082	1.06	0.79, 1.21	0.6	-3.25	-0.89	1.13	3.85
Sserve136	-0.057	0.082	1.03	0.79, 1.21	0.3	-3.50	-1.14	0.88	3.60
Sserve137	-0.225	0.082	0.90	0.80, 1.20	-0.9	-3.66	-1.30	0.72	3.45
Sserve138	-0.481	0.082	1.01	0.80, 1.20	0.2	-3.90	-1.54	0.48	3.20
Sserve139	0.161	0.082	1.02	0.80, 1.20	0.2	-3.27	-0.91	1.11	3.83
Sserve140	-0.455	0.082	0.91	0.80, 1.20	-0.8	-3.88	-1.52	0.50	3.22
Sserve141	0.146	0.082	0.80	0.80, 1.20	-2.0	-3.33	-0.95	1.07	3.78
Sserve142	0.026	0.082	1.01	0.80, 1.20	0.2	-3.45	-1.06	0.95	3.66
Sserve143	0.161	0.245	0.80	0.80, 1.20	-2.0	-3.31	-0.93	1.09	3.80

Note. *Constrained, significance level = 0.000

(*continued*)

Table 3. (*continued*)

Variable			Weighted Fit						
Student academic culture	Estimates	Error	MNSQR	CI	t	Item thresholds			
Use of time									
Sacad144	0.103	0.082	0.92	0.78, 1.22	-0.7	-2.17	-1.26	0.69	3.14
Sacad145	0.077	0.082	1.01	0.78, 1.22	0.1	-2.20	-1.29	0.66	3.12
Sacad146	0.387	0.081	1.05	0.78, 1.22	0.5	-1.89	-0.98	0.97	3.42
Sacad147	-0.567*	0.141	1.01	0.79, 1.21	0.1	-2.84	-1.93	0.02	2.47
Student academic behavior									
Sacad148	-0.676	0.065	0.73	0.77, 1.23	-2.6	-2.66	-1.38	-0.38	1.72
Sacad149	-0.932	0.067	0.84	0.77, 1.23	-1.4	-2.92	-1.63	-0.64	1.46
Sacad150	-0.482	0.064	0.81	0.78, 1.22	-1.8	-2.47	-1.18	-0.19	1.91
Sacad151	-0.147	0.063	0.77	0.79, 1.21	-2.3	-2.13	-0.85	0.15	2.24
Sacad152	0.358	0.061	1.04	0.81, 1.19	0.5	-1.63	-0.34	0.65	2.75
Sacad153	-0.300	0.064	1.16	0.78, 1.22	1.4	-2.29	-1.00	-0.00	2.09
Sacad154	1.086	0.060	1.09	0.82, 1.18	0.9	-0.90	0.38	1.38	3.48
Sacad155	1.093	0.168	1.27	0.82, 1.18	2.7	-0.89	0.39	1.39	3.48

culture of student learning on campus). The division of this service scale into three sub-scales of measurement is determined based on the results of item validation tests using principal component analysis techniques. Then, to determine whether each sub-scale is really supported by items that have adequate fit statistics, then it is processed again by applying the Rasch model technique. As previously explained, the analysis using the Rach model technique was carried out using Conquest 2.0 software to calculate the fit statistics and thresholds values for each item of these variables. The results of this model test are expected to help the team to decide whether each of the measurement scales developed for these variables meet the "unidimensional" principle.

With respect to all of the validated items, proving that all of the items "university administration services" and "student academic culture" can be classified as acceptable instrument items because they have MNSQR scores in the range of 0.5–1.5 or are considered to meet productive standards items. The estimation results with the Rasch model are reported in Table 3.

In order for the analysed items to be more meaningful to the readers, each item needs to be described as presented in the list of questions of the instrument as in Table 4.

Examining fit statistics of t values with the exception items "Sserve120" and "Sserve128", all other items do not exceed t values 3. This means in average they have adequate predictive ability. So it can be interpreted that behaviours or attitudes being measured are able to predict. The three items that obtained $t \geq 3$ were categorized as not having adequate and reasonable predictability. This explanation is relevant with t interpretation guidelines. However, it should be noted that the two problematic items do not automatically invalidate the function of each item as an indicator of the latent variable representing them. Moreover, the items were tested through a fairly large sample that could increase t values obtained (Wu & Adams, 2007). Therefore, instead of using this

Table 4. Scale items of administrative services and student academic culture

Scale/Item	Questions
Staff attitude	
Sserve116	Employees smile while serving students
Sserve117	Make eye contact while looking at the student
Sserve118	Mention the name of students while providing services
Sserve119	Listening to students' complaints while nodding the head
Sserve120	Ask open-ended questions such as "can I help you?"
Sserve121	Show empathy when handling student complaints/problem
Sserve123	Give reasons before saying disagree
Sserve124	Admit unintentional mistakes
Sserve129	Try to help solve student problems
Sserve131	End the conversation with a positive note
Direct communication	
Sserve122	Stand quietly when talking to you
Sserve125	Look attentively when dealing with you
Sserve126	Concentrate on what you are talking about/avoid other interventions
Sserve127	Sounds "emm", "yes" and nodding the head
Sserve128	Ask open-ended questions and ready to help if needed
Sserve130	Employees stay calm so they can provide services thoughtfully
Sserve132	Not to interrupt when talking to students
Sserve133	Summarize what customers said at the end of the meeting
Telephone communication	
Sserve134	Answer your phone calls as quickly as possible
Sserve135	Answering your phone impressed with a smile and friendly
Sserve136	Talk on the phone in a low but clear tone
Sserve137	Answer by mentioning the person's name, or: "what can I help you with"
Sserve138	Asking open-ended questions what students want
Sserve139	Trying to record what is said when you call
Sserve140	When called, trying to help students need information
Sserve141	Warm up the conversation before hanging up the phone
Sserve142	End the call making sure to follow-up your request
Sserve143	Tell her name that if at any time you call again
Use of time	

(*continued*)

Table 4. (*continued*)

Scale/Item	Questions
Sacad144	Use my free time to study
Sacad145	I do my assignments even though the lecturers do not set a deadline time
Sacad146	I like going to the library instead of wasting available time
Sacad147	Trying to find fellow students to discuss lectures
Student academic behavior	
Sacad148	Prefers courses that prioritize a lot of quality literature
Sacad149	Prefers lectures that use a variety of trusted sources
Sacad150	Prefers lectures that use international and local literature
Sacad151	Prefers lectures that only use local literature
Sacad152	I prefer lectures with a little literature
Sacad153	I prefer lecturers who give less assignments
Sacad154	I prefer lecturers who rarely come to the class
Sacad155	I like students who always don't focus on academics

procedure as the only tool for selecting items, it is more appropriate if it is used as a consideration to detect which items should be considered problematic, and if necessary use its findings as fair information to be considered in developing items. In addition to other quality categories (MNSQ, item delta, thresholds, discrimination index), the two items still obtained an adequate fit. The final result of the fit t statistic may be able to explain truth about different behaviours shown by each item in the measurement process. It is likely that each item will not perform the same when describing their representative latent variables.

Furthermore, based on the criteria of statistical values of item thresholds, it indicates that all items as in Table 3 and Table 4 do not have an inconsistency problem in the order of the answer choice categories for each item (disordered item thresholds). The results of the technical analysis of the Rasch model show that the university administration service items and student academic culture in each sub-scale of measurement meet the interpretable item responses. The consistency of the distance per response category also indicates that each item response category has received monotonous answers from the respondents. In addition, the observation results show that the thresholds of the items are in line with the chronology of the difficulty level of the answer desired by the research measurement model proposed in this study.

4 Conclusion

To conclude, it can be concluded that this instrument is able to reveal the actual state of attitude measured in the field through a Likert scale of types of university organizational culture in the field. Based on the values of the item discrimination ability index achieved,

it shows that generally each item has a significant relationship with the total response score for all sub-scales of measurement, namely achieving a discrimination index >0.4, with a significance level of <0.01. In fact, there are a number of items that have a discrimination index value of 0.85, namely "Sserve121" and with a value of 0.84, namely "Sserve 141" and "Sserve 143" items. In the IRT model, it is recommended to choose items that have an index value > 0.4 (Wu & Adams, 2007: 64). The results of the item analysis in Table 11.3 and Table 11.4 show that the items in the measurement scale of "university administration services" and "student academic culture" obtained indexes in the range 0.47–0.85. These results indicate that all items have high quality in terms of the ability to distinguish respondents according to the level of individual autonomy. The last one as listed below the separation reliability index table for each sub-scale of measurement has reached a high reliability index, which is in the range of 0.776–0.991. The index achievement shows that most of the variance proportions for each subscale are correct.

References

Ahmad, T. (2015). Preparing for the future of higher education. *On the Horizon, 23*(4), 323-330. https://doi.org/10.1108/OTH-06-2015-0029

Basir, S. A., Davies, J., Douglas, J., & Douglas, A. (2017). The influence of academic culture on quality management system ISO 9001 maintenance within Malaysian universities. *Journal of Higher Education Policy and Management, 39*(3), 320-340. https://doi.org/10.1080/1360080X.2017.1298199

Burhanuddin, B. (2017). Leading The University To Build A Greater Performance. *KnE Social Sciences, 1*(3). https://doi.org/10.18502/kss.v1i3.724

Burhanuddin, B., & Sunarni, S. (2016). *The development of measurement model for organizational culture and its effects on leadership effectiveness at Primary Schools in the City of Malang.* Retrieved from Malang:

Cramer, D. (2003). *Advanced quantitative data analysis.* Maidenhead: Open University Press.

Creswell, J. W. (2005). *Educational research: planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, New Jersey: Pearson Merrill Prentice Hall.

Creswell, J. W. (2014). *Research design. Qualitative, quantitative, and mixed methods approaches.* Sage: Los Angeles.

Gray, D. E. (2009). *Doing research in the real world* (2nd ed.). Los Angeles: Sage.

Hair_Jr., J. F., Black, W. C., Babin, B. J., & Rolph E, A. (2016). *Multi variate data analysis.* Upper Saddle River, New Jersey: Prentice Hall.

Jones, S., Lefoe, G., Harvey, M., & Ryland, K. (2012). Distributed leadership: A collaborative framework for academics, executives and professionals in higher education. *Journal of Higher Education Policy and Management, 34*(1), 67-78.

Kelly, A. E., Baek, J. Y., Lesh, R. A., & Bannan-Ritland, B. (2008). Enabling inovations in education and systematizing their impact. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education: Innovations in science. technology, engineering, and mathematics learning and teaching.* New York: Routledge Taylor & Francis Group.

Levin, R. I., Rastogi, S., Rubin, D. S., & Siddiqui, M. H. (2014). *Statistics for management*. NOIDA, India: Pearson.

Linacre, J. M. (2009). *A user's guide to WINSTEPS MINISTEP: Rasch-model computer programs.* Chicago, IL.

Richey, R. C., Klein, J. D., & Nelson, W. A. (2004). Developmental research: Studies of instructional design and development. In D. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2 ed., pp. 1099–1130). Mahwah, N.J: Lwrence Erlbaum Associates, inc.

Smith, G., Smith, A., & Clarke, A. (2007). Evaluating service quality in universities: a service department perspective. *Quality Assurance in Education, 15*(3), 334-351. https://doi.org/10.1108/09684880710773200

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5 ed.). New York: Routledge Taylor & Francis Group.

Wu, M. L., & Adams, R. J. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach.* Melbourne: Educational Measurement Solutions.

Wu, M. L., Adams, R. J., & Haldane, S. A. (2005). ConQuest 2.0: Australian Council for Educational Research University of California, Berkeley.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest Version 2.0.* Victoria, Australia: Acer Press.