



Classical Test Theory: Measurement for Fashion Design Double Track High School Program

*I Russanti¹, Eko Hariadi², T Rijanto³

¹Department of Vocational Education, Sekolah Pascasarjana, Universitas Negeri Surabaya, Surabaya 60231, Indonesia

Article Info

Article history:

Received August 1, 2022

Revised August 22, 2022

Accepted September 7, 2022

Available Online November 30, 2022

Keywords:

Classical Test

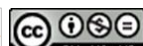
Double Track

Fashion Design

High School

ABSTRACT

The Double Track High School Program is a program that provides the provisions of high school students in order to have competencies such as vocational students, one of which is the Fashion Design competency. Tests are needed to measure students' cognitive abilities. This research is a test that aims to obtain information on the validation of the Fashion Design cognitive test items using the classical test theory. This quantitative descriptive study uses a multiple choices test with four choices of 40 items. Participants' responses were collected through Google form for Fashion Design class of 2019 and 2018 totaling 122 respondents. The analysis technique with the strata program was analyzed using classical test theory. Based on the analysis of the cognitive validation tests of fashion design competencies Double track high school program results obtained the validation of the test uses classical test theory with a test reliability coefficient of 0.5729 for the number of tests as many as 40 questions, meaning the question items has moderate reliability. The conclusion is a test can be used for fashion design double track high school program.

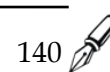


<https://doi.org/10.46627/sipose>

INTRODUCTION

Double Track High School Program is a program that provides provisions for high school students to have competences like vocational students. The goal is to solve the problem for high school graduates who are unable to continue their education to tertiary institutions. On average they choose to work, but because they do not have the skills like Vocational High School (VHS) graduates, many find it difficult to get a job. The Double Track concept to face the global battle and fill the workforce in the industry requires a link and match between education and employment vacancies.

The Double Track Senior High School (SHS) program is almost similar to the regular track, theoretical deepening is given to students from grade X-XII. The difference is that in the Double Track SHS program, students can choose special training in their area of expertise in the second semester of Class XII. A competency certificate will be given upon graduation which can be used as a portfolio when they are looking for work. The skills provided to students will provide mental readiness and skills to work. Double track high school graduates are more confident because they have competency certificates. One of the areas of expertise at Double Track High School is Fashion Design. In this field of expertise, it provides provisions for students to start making designs, to transforming them into clothing. Hence, to make a comprehensive value of the fashion class it is needs to assess with a reliable measurement test as an assessment tool. Assessment is an important aspect of learning to measure students' cognitive abilities, academic and intellectual



skills. Assessment development using certain techniques to know student learning outcomes One of them uses tests.

To be able to measure the mastery of cognitive abilities of students in the field of expertise of fashion design in high school double track is done using tests. A test that is common or often used to measure the ability of students in schools is a test instrument. The test instrument is considered capable of accommodating any material that has been taught by the teacher to students in learning in the form of questions (Susdelina, 2018). The quality of a test kit can be seen by conducting quantitative analysis using an approach, namely the classical test theory technique (Karim et al., 2022; Ramadhan et al., 2020; Yulisharyasti et al., 2023). Several aspects that are considered in classical test theory are the level of difficulty of the items, the distinguishing power of distribution of answer choices, and reliability of test scores. Classical test theory has been used for decades to determine reliability and other characteristics of measurement instruments Crocker. Item response theory is a modern measurement theory commonly used in item analysis. In this theory, a mathematical model is used to connect the characteristics of the items with the responsiveness of the item. This relationship is represented by the grain characteristic curve. The test instrument is said to have good quality if it has high validity and reliability. The higher the validity and reliability of an instrument, the more accurate the data obtained from a study (Gao et al., 2020; Syakur et al., 2019). To understand the abilities of highly intelligent students through tests, a central tendency measurement is carried out, including scale scores and ratings, this is the way teachers understand student abilities (Lebeau et al., 2020). Through the evaluation of learning outcomes tests is one of the most effective ways to improve the teaching and learning process (Dwijayani, 2019; Hamilton et al., 2021; Pei & Wu, 2019; Supena et al., 2021). Processing of learning outcomes tests in order to improve the teaching and learning process, one of which is to do item analysis (Albus et al., 2021; Wayan Mertha & Mahfud, 2022; Yilmaz İnce et al., 2020).

Hence, the novelties of this research are contributes to the analysis of the classical test in the field of fashion design at Double Track High School, to give information on the validity of test items using classical tests for teachers, and to became the literature on fashion design field test trials for cognitive abilities. The research objectives of this study is to obtain validation information on cognitive test items Fashion Design SHS Double Track using classical test theory.

RESEARCH METHOD

This type of descriptive research with a quantitative approach to determine test trials which aims to obtain validation information on cognitive test items Fashion Design High School Double Track using classical test theory. The test used is multiple choice with four choices of 40 items. Participant responses were collected via Google form with 122 respondents of Fashion Design students 2019 and 2018 in 2020. The fashion design test instrument for SHS Double Track contains cognitive abilities which consist as in Table 1.

Table 1. Item numbers.

Materials	Item Numbers	Total
Communication Studies	1-4	5
Excellent Service	5-12	5
Design	13-27	15
Textile	28-32	5
Construction	33-36	5
Sewing Technique	37-40	5

Based on Table 1 there are 6 competencies measured on the test instrument, namely: communication science, excellent service, design, textiles, construction and sewing techniques with the number of distributions for cognitive abilities in each competency is 8.00% except Design is 24.00% because it consists of competences about elements and principles of design, proportion and fashion drawing.



The analysis technique uses the Kuder Richardson (KR-20) method state program which is analyzed using classical test theory. Furthermore, the calculation results will be interpreted against the coefficient or value of r as follows: 1) If r_{11} is equal to or greater than 0.70, then the test results being tested for reliability have high reliability. 2) If r_{11} is less than 0.70, then the test results that are being tested for reliability are declared not to have high reliability. To interpret the item difficulty index, the classification is used, namely: 1) Problems with P 0.00 to 0.29 are difficult questions, 2) Problems with P 0.30 to 0.69 are medium questions, 3) Problems with P 0.70 to 1.00 are easy questions (Fatimah et al., 2020; Nadhifah & Jauhariyah, 2021; Rustana et al., 2021; Sari & Oktavia, 2023). Questions that are considered good are questions that are in the medium category, namely questions that have a difficulty index of 0.30 to 0.69 (Journal, 2013; Pargito, 2021; Rahmah et al., 2020).

RESULTS AND DISCUSSION

This section describes the data obtained from the results of the test questions that have been carried out as many as 122 google form respondents then identified through the stata program with the Kuder-Richarson KR-20 method the results are obtained according to Figure 1.

. kr20 var1-var40				
Kuder-Richarson coefficient of reliability (KR-20)				
Number of items in the scale = 40				
Number of complete observations = 122				
Item	Obs	Item difficulty	Item variance	Item-rest correlation
var1	122	0.9754	0.0240	0.1199
var2	122	0.4508	0.2476	-0.1278
var3	122	0.3525	0.2282	0.1313
var4	122	0.9836	0.0161	0.3647
var5	122	0.5328	0.2489	0.1177
var6	122	0.7049	0.2080	0.2237
var7	122	0.0492	0.0468	0.0193
var8	122	0.8770	0.1078	0.2005
var9	122	0.8197	0.1478	0.0539
var10	122	0.5656	0.2457	0.1487
var11	122	0.6639	0.2231	0.2103
var12	122	0.6230	0.2349	0.0013
var13	122	0.8934	0.0952	0.2063
var14	122	0.1148	0.1016	-0.0400
var15	122	0.0410	0.0393	0.0284
var16	122	0.1475	0.1258	0.2020
var17	122	0.9180	0.0752	0.2001
var18	122	0.5492	0.2476	0.3033
var19	122	0.2377	0.1812	0.1040
var20	122	0.8607	0.1199	0.1170
var21	122	0.7951	0.1629	0.2370
var22	122	0.9180	0.0752	0.2169
var23	122	0.0410	0.0393	0.1191
var24	122	0.7049	0.2080	0.1260
var25	122	0.9180	0.0752	0.2761
var26	122	0.8770	0.1078	0.2218
var27	122	0.8443	0.1315	0.2402
var28	122	0.7049	0.2080	0.1412
var29	122	0.0654	0.0413	-0.1108
var30	122	0.4918	0.2499	0.1905
var31	122	0.6393	0.2306	0.1838
var32	122	0.5000	0.2500	0.0842
var33	122	0.4918	0.2499	0.2576
var34	122	0.9344	0.0613	0.3174
var35	122	0.8934	0.0952	0.1764
var36	122	0.9344	0.0613	0.0219
var37	122	0.6557	0.2257	0.2565
var38	122	0.2131	0.1677	-0.0687
var39	122	0.1311	0.1139	-0.0113
var40	122	0.6230	0.2349	0.1868
Test		0.5934		0.1412
KR20 coefficient is 0.5729				

Figure 1. KR 20.

Table 2. Sample of table.

Competency	Item Number	%	Low Level	Moderate Level	High Level
Communication Studies	1-5	8.00%	1, 4	2, 3, 5	
Excellent Service	6-10	8.00%	6, 8, 9	10	7
Design	11-25	24.00%	13, 17, 30, 24, 25	11, 12, 18, 21, 22	14, 15, 16, 19, 23
Textile	26-30	8.00%	26, 27, 28	29	30
Construction	31-35	8.00%	34, 35	31, 32, 33	
Sewing Technique	36-40	8.00%	36	37, 40	38, 39

Based on the Table 2, the lowest level of problem difficulty is 0.04 and the highest is 0.98. If the value range is 0-0,299 it means that the question is very difficult, namely on item number 7, 14, 15, 16, 19, 23, 30, 38, 39. For the value range of 0.70-1.00 the difficulty level of the easy questions is on the items number 1, 4, 6, 8, 9, 13, 17, 20, 22, 24, 25, 26, 27, 28, 34, 35, 36. While the rest is in the range 0.30-0.69, meaning that the questions are medium. Kr20 results in Figure 1 indicate that the coefficient of the items is 0.57, meaning that the test results being tested for reliability are



declared to have moderate reliability. A test's reliability has crucial implications for the quality of decisions that are made on the basis of an individual's test scores. Recall that we can never know an individual's "true" level on an unobservable psychological construct (Pargito, 2021; Rahmah et al., 2020).

Based on Table 1, 9 items are at the high difficulty level, 17 items are at the low difficulty level and the remaining 14 items are at the medium difficulty level. For the distribution of the level of difficulty of the questions can be analyzed in Table 2. In the communication science competency, the number of questions is 8.00% of the total 5 items which are at the low and medium difficulty level, meaning that the questions are relatively easy to moderate. At excellent service competence with a number of questions as much as 8.00% with a total of 5 items spread at a level of difficulty starting from relatively low, sufficient and high. This competency is a core competency in the field of fashion design expertise so that the number of questions is 24.00% with a total of 15 items. This competency includes the competence of applying design elements and principles, drawing fashion and proportions. The distribution of the level of difficulty is evenly distributed, starting from relatively low, medium to high. In textile competency, the number of questions was 8% with a total of 5 items. The item difficulty level spreads from relatively low, medium to high, meaning that the difficulty level of the questions spreads from relatively easy to high. In sewing competency, the number of questions was 8.00% with a total of 5 items. Low to medium difficulty level means the questions are relatively easy to moderate. In the construction competency pattern, the number of questions is 8.00% with a number of 5 items. The difficulty level of the questions is evenly distributed from low, medium to high. This means that the level of difficulty of the questions is evenly distributed from relatively easy to difficult.

Table 1 shows that the KR-20 result is 0.57, which means that the question has medium reality. The aspects that really determine the quality of the items are the difficulty level and the distinguishing power of the questions (Cakranegara et al., 2022; Secules et al., 2021; Spinelli et al., 2020; Zednik, 2021). The quality of a measurement instrument is expressed in the reliability and validity of the scores collected by this instrument. Scale validity refers to the extent of the correspondence between the variation in the scores on the test and the variations among respondents on the tested constructs (Himelfarb, 2019). Thus, the results of research on instrument validation, questions can be used to measure cognitive abilities in the field of expertise Fashion Design SHS Double Track because it has medium reality.

CONCLUSION

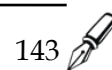
Based on the validation analysis of the cognitive test of fashion competency in the design of the Double Track Senior High School program, the results obtained that the test validation uses the classical test theory with a test reliability efficiency stated to be quite reliable. Furthermore, the KR-20 result is stated that the question has medium reality. Hence, through this reliability and validity, the test can be used for fashion design double track high school program. Moreover, this research limitation is the subject of the research are limit to one Fashion Design High School Double Track, and so further research may can conduct research to other Fashion Design High School Double Track to detailed and compare the rest results.

ACKNOWLEDGEMENTS

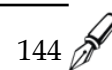
The authors thank to Sekolah Pascasarjana Universitas Negeri Surabaya that gives extraordinary supports. They also thank many stakeholders involved in the research completion.

REFERENCES

- Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers and Education*, 166, 1-10. <https://doi.org/10.1016/j.compedu.2021.104154>
- Cakranegara, P. A., Kurniadi, W., Sampe, F., Pangemanan, J., & Yusuf, M. (2022). The impact of goods product pricing strategies on consumer purchasing power: A review of the literature.



- Jurnal Ekonomi*, 11(03), 1-10. <https://doi.org/10.3389/fpsyg.2021.720151>
- Dwijayani, N. M. (2019). Development of circle learning media to improve student learning outcomes. *Journal of Physics: Conference Series*, 1321(2), 1-5. <https://doi.org/10.1088/1742-6596/1321/2/022099>
- Fatimah, S., Elzamzami, A. B., & Slamet, J. (2020). Item analysis of final test for the 9th grade students of SMPN 44 surabaya in the academic year of 2019/2020. *JournEEL (Journal of English Education and Literature)*, 2(1), 34-46. <https://doi.org/10.51836/journeel.v2i1.81>
- Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education*, 7(1), 1-18. <https://doi.org/10.1186/s40594-020-00225-4>
- Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. In *Journal of Computers in Education*, 8(1), 1-10. <https://doi.org/10.1007/s40692-020-00169-2>
- Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, 33(2), 151-163. <https://doi.org/10.7899%2FJCE-18-22>
- Karim, A., Syamsuddin, I., Jumarding, A., & Amrullah, A. (2022). The effect of gender independence and leadership style on audit quality in makassar public accounting offices. *International Journal of Social Science Research and Review*, 5(7), 114-126. <https://doi.org/10.47814/ijssrr.v5i7.341>
- LeBeau, B., Assouline, S. G., Mahatmya, D., & Lupkowski-Shoplik, A. (2020). Differentiating among high-achieving learners: A comparison of classical test theory and item response theory on above-level testing. *Gifted Child Quarterly*, 64(3), 219-237. <https://doi.org/10.1177/0016986220924050>
- Nadhifah, A., & Jauhariyah, M. N. R. (2021). The use of the quizizz application in the scientific literacy assessment of high school students on global warming material. *Jurnal Ilmiah Pendidikan Fisika*, 5(2), 162. <https://doi.org/10.20527/jipf.v5i2.3309>
- Pargito, P. (2021). Development of assessment instruments based on higher order thinking skills in thematic learning. *The Internasional Journal Social Sciences*, 3(1), 165-175.
- Pei, L., & Wu, H. (2019). Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. *Medical Education Online*, 24(1), 1-10. <https://doi.org/10.1080/10872981.2019.1666538>
- Rahmah, N., Yusrizal, Y., & Syukri, M. (2020). Analysis of multiple-choice question (MCQ) of physics final examination in senior high school. *Journal of Physics: Conference Series*, 1460(1), 1-4. <https://doi.org/10.1088/1742-6596/1460/1/012143>
- Ramadhan, S., Sumiharsono, R., Mardapi, D., & Prasetyo, Z. K. (2020). The quality of test instruments constructed by teachers in bima regency, indonesia: Document analysis. *International Journal of Instruction*, 13(2), 507-518. <https://doi.org/10.29333/iji.2020.13235a>
- Rustana, C. E., Aminah, S., & Budi, A. S. (2021). The development of harmonic oscillation e-module based on problem based learning (pbl) for helping improvement of students' higher order thinking skills (HOTs). *Journal of Physics: Conference Series*, 1869(1), 1-5. <https://doi.org/10.1088/1742-6596/1869/1/012174>
- Sari, A. P., & Oktavia, W. (2023). Students' difficulties with parallel structure in the structure and written expression section in TOEFL test. *Journal of English Language Teaching English*, 12(1), 106-121. <https://doi.org/10.24036/jelt.v12i1.121498>
- Secules, S., McCall, C., Mejia, J. A., Beebe, C., Masters, A. S., L. Sánchez-Peña, M., & Svyantek, M. (2021). Positionality practices and dimensions of impact on equity research: A collaborative inquiry and call to the community. *Journal of Engineering Education*, 110(1), 19-43. <https://doi.org/10.1002/jee.20377>
- Spinelli, M., Lionetti, F., Pastore, M., & Fasolo, M. (2020). Parents' stress and children's psychological problems in families facing the COVID-19 outbreak in Italy. *Frontiers in*



- Psychology*, 11, 1–7. <https://doi.org/10.3389/fpsyg.2020.01713>
- Supena, I., Darmuki, A., & Hariyadi, A. (2021). The influence of 4C (constructive, critical, creativity, collaborative) learning model on students' learning outcomes. *International Journal of Instruction*, 14(3), 873–892. <https://doi.org/10.29333/iji.2021.14351a>
- Syakur, A., Junining, E., & Sabat, Y. (2019). Application of E-learning as a method in educational model to increase the TOEFL score in higher education. *Journal of Development Research*, 3(2), 111–116. <https://doi.org/10.28926/jdr.v3i2.88>
- Wayan, M. I., & Mahfud, M. (2022). History learning based on wordwall applications to improve student learning results class X IPS in MA as'adiyah ketapang. *International Journal of Educational Review, Law And Social Sciences (IJERLAS)*, 2(5), 507–612. <https://doi.org/10.54443/ijerlas.v2i5.369>
- Yılmaz İnce, E., Kabul, A., & Diler, İ. (2020). Distance education in higher education in the COVID-19 pandemic process: A case of isparta applied sciences university. *International Journal of Technology in Education and Science*, 4(4), 343–351. <https://doi.org/10.46328/ijtes.v4i4.112>
- Yulisharyasti, L., Nurdin, A., Aulia, N., Arfa, F. A. H., & Fadjryani. (2023). Analyzing the quality of measurement instruments of multiple choice questions on class XI economics material in public high school 3 gorontalo through classical test theory and rasch models. *Parameter: Journal of Statistics*, 3(1), 28–38. <https://doi.org/10.22487/27765660.2023.v3.i1.16417>
- Zednik, C. (2021). Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy and Technology*, 34(2), 265–288. <https://doi.org/10.1007/s13347-019-00382-7>
-

Author (s):

* Irma Russanti (Corresponding Author)
Department of Vocational Education Sekolah Pascasarjana
Universitas Negeri Surabaya,
Jl. Lidah Wetan, Surabaya 60213, Indonesia
Email: irmarussanti@unesa.ac.id

Eko Hariadi
Department of Vocational Education Sekolah Pascasarjana
Universitas Negeri Surabaya,
Jl. Lidah Wetan, Surabaya 60213, Indonesia
Email: ekohariadi@unesa.ac.id

Tri Rijanto
Department of Vocational Education Sekolah Pascasarjana
Universitas Negeri Surabaya,
Jl. Lidah Wetan, Surabaya 60213, Indonesia
Email: tririjanto@unesa.ac.id

