



Quality Evaluation of Try Out Competency Test of Indonesian Medical Record and Health Information Association to Support Graduation Achievement

1st Lily Widjaja,

Prodi D-3 RMIK, Universitas Esa Unggul,
Jakarta, Indonesia,
liliwidjajaiwan@gmail.com

2nd Kori Puspita Ningsih,

Prodi D-3 RMIK, Universitas Jenderal Achmad Yani
Yogyakarta,
Yogyakarta, Indonesia,
puspitakori@gmail.com

3rd Ida Sugiarti,

Prodi D-3 RMIK, Poltekkes Kemenkes Tasikmalaya,
Jawa Barat, Indonesia,
Sugiarti.ida@gmail.com

ABSTRACT

The try out of APTIRMIKI's competency test was prepared and developed according to PMIK's blueprint. The expected output is a description of students' ability to master the competencies tested. The competency test try-out cannot measure students' abilities appropriately if the try-out question does not have good quality so the goal of the try-out competency test is not maximally reached. The purpose of this study was to test the quality of the questions about the try out of the competency test by APTIRMIKI in 2018. The type of this research is descriptive quantitative. The population of this study was all final year students from 38 DIK RM3 educational institutions totaling 2288 participants who tried out the competency test. The sample in this study is a saturated sample in which all members of the population are included as samples. Data collection techniques with the method of documentation and interviews. Data analysis techniques using the application of instrument analysis problems. The analysis was conducted to determine the validity, reliability, level of difficulty, item validity, reliability, distinguishing features, quality of deceivers. The results of the analysis of the items: valid 43,33% (78) and invalid 56,67% (102). The reliability value of the try and competency test is 0.45. A total of 97 items (53,88%) need to be revised. A total of 48.33% (87) questions were at a moderate difficulty level. Distractors function well in terms of as many as 36,11% (65) while needing to be revised by 48,89% (88). It can be concluded that the quality of the APTIRMIKI competency test tryout is not good enough so that it needs to be improved.

Keywords: *quality of questions, tryout, competency tests.*

I. INTRODUCTION

Competency test is a process of measuring the knowledge, skills, and training of students in tertiary institutions organizing health programs, especially the Medical Record Study Program & Health Information Management (RMIK). To prepare students or candidates who will submit competency tests at the end of the study period, APTIRMIKI (Indonesian Medical Record and Health Information Management Association) has conducted competency trials (Permenkes, 2013).

Organizations try competency tests that need to be evaluated, especially identifying, refining and improving the quality of the questions used.. Competency trials that play an important role in competency tests, so that competency tests are needed must also be per good competency test standards and can be made to measure the competency of Health Information Medical Recorders (PMIK) competencies (Akbar Rizal, 2018). By discussing the item analysis, information about the item used will be obtained (Srika N.P et al, 2018). Evaluation through analysis of each item is carried out in the hope of finding a variety of information, which is feedback to make improvements, improvements, and refinement of the items so that in the future compiled or designed in the future. the team developer can measure what



is obtained that is carried out continuously and carried out by APTIRMIKI periodically, completely, transparently, and systematically, to obtain competency standards per PMIK (Arifin, 2017).

Good or quality approved instruments, agreed to satisfy tests, among other things; (1) valid; (2) reliable; (3) level of difficulty of items; (4) grain distinguishing power; (5) distributing answer choices (for objective form questions) or checking each answer option. An instrument that cannot produce conclusions that do not correspond to reality. The items in the test reflect the importance of student competence, so the level of difficulty, differentiation, validity & reliability of the test items test questions must be ensured to produce good output (Afiyanti, 2008; Iskandar, Akhbar; Rizal, 2017).

The test of competency testing is prepared and developed according to the PMIK blueprint. The expected output from the competency trial is a description of students' abilities in competency competencies tested in the comet test. The tryout test competency test cannot measure the ability of students to properly question. The tryout question does not have good quality with the aim of tryout.

II. METHODS

To describe the quality of the try out APTIRMIKI's competency test used a descriptive qualitative research. The quantitative approach used in the analysis of question quality empirically examines student answer items in the TOUKOM application database. The analysis carried out includes item validity, reliability, distinguishing features, level of difficulty, quality of deceivers.

The location of this study is the educational institutions of APTIRMIKI members throughout Indonesia, amounting to 42 educational institutions who participated in

the competency test and the time of the study was May to December 2019. The population of this study was all final year students from 42 D3 RMIK educational institutions, totaling 2288 participants try out the competency test. The sample in this study is a saturated sample in which all members of the population are included as samples.

Data collection was carried out using the documentation method and the interview method. The documentation method is used to collect data by looking at the documentation materials that exist in the CBT Try Out APTIRMIKI Competency Test application; in the form of an electronic question sheet, the results of the answer along with the answer key, and a list of test participants try out. The data is obtained directly from the APTIRMIKI R&D archive. As for the interviews conducted on the compiler try out questions that aim to obtain verbal data research problems. Data analysis was performed using the question analysis instrument application. The analysis was conducted to determine the validity, reliability, level of difficulty, item validity, reliability, distinguishing features, quality of deceivers.

III. RESULT AND DISCUSSION

A. Validity

Items that were proven to be invalid on the try out questions because it could cause the test to see student competencies not being optimal. Based on the analysis of the APTIRMIKI competency test, it is known from the number of questions, in the 2018 package questions about 78 items (43.33%) are valid questions, while 102 items (56.67%) are invalid questions (table 1). Problems that can be reused are problems that must be corrected or asked for invalid questions.

Table 1 Item Validity Test Result

Item validity test		
Criteria	Amount	Percentage
Valid	78	43,33
Invalid	102	56,67
Amount	180	100

Analysis of test quality is a stage that must be taken to determine the degree of quality of a test, both as a whole and the items that are part of the test (Oluwatoyo, 2012). The test is said to have good quality if it fulfills two things, namely accuracy or validity and determination or reliability and the test itself must meet the internal characteristics of the quality of the questions in terms of qualitative and quantitative. The quality of the questions in terms of qualitative lies in the construction, terms of language and test material. The qualitative analysis of the implementation uses the rules contained in writing the test and is carried out before the test is used or tested. The quality of the questions seen in quantitative terms that include the level of difficulty, distinguishing features, the spread of answers and the reliability of the test (Istika N et al, 2019)

B. Reliability

The results of the try out value analysis of the APTIRMIKI competency test with the help of the application generated reliability values. A test is said to be reliable if it always gives the same results when tested in the same group at different times or opportunities (Rizal, 2017). The concept of reliability in terms of the reliability of measuring instruments is closely related to the problem of measurement error (Salim et al., 2016). The measurement error itself shows the extent to which the inconsistency of measurement results

occurs when repeated measurements of the same group of subjects. While the concept of reliability in the sense of the reliability of the measurement results is closely related to the error in sampling that refers to the inconsistency of the measurement results if the measurements are repeated on different groups. This is similar to what was expressed (Solichin, 2017) states that the reliability of the assessment tool is the accuracy or the severity of the tool in assessing what is valued. That is, whenever the assessment tool is used will give relatively the same results. A test is said to be reliable if the test is given repeatedly to give the same results (Istika N et al, 2019).

The results of the APTIRMIKI competency test try out the value based on the application assistance produced the reliability value of the tryout competency test is 0.45; so included in the category enough. According to Basuki and Hariyanto (2014: 119), if the reliability calculation is between the coefficients of 0.00 - 0.19, it is included in the reliability in the very low category. The results of the calculation of reliability are between the coefficients 0.20 to 0.39 then included in the reliability in the low category. The calculation of the reliability of the competency test is between the coefficients 0.40 - 0.69 then it is included in the reliability in the sufficient category. reliability calculation is between the coefficients 0.70 - 0.89 it is included in the reliability in the high category. The results of the calculation of reliability are between the coefficients 0.90 - 1.00 then included in the reliability in the very high category. The higher the reliability index (constancy/accuracy) of a test, the higher the level of validity (Setyaningrum.P.M.P, et al, 2018). This is in line with research (Amalia & Widayati, 2012) stating that reliability as a supporter of item validity



is formed. The high and low-reliability coefficient is influenced by several factors including: (1) the length of the test, (2) the speed (time limit) doing the test, (3) the homogeneity of the group, (4) the difficulty of the items (Hanifah, 2014).

C. Distinguishing power

Distinguishing power is the ability of an item to distinguish between the try out participants who master the material in question and the try out participants who have not mastered the material being tested. The difference in the power of items is an index that shows the level of ability of items to distinguish high-achieving groups (upper groups) from low performing groups (lower groups) among test takers. The low power difference is usually caused by the level of functioning of the deceit items, besides the deception also has an impact on the level of difficulty of the items because if there are one or two distractors on a non-functional item then the index of the difficulty level of the items will decrease because the chances of the test takers to answer correctly increasingly. The power of distinguishing questions is the ability of questions to distinguish between high-ability students and low-ability students (Daryanto, 2010). The power of distinguishing questions is the ability of the questions with their scores to distinguish test participants from high and low groups (Hanifah, 2014). In other words, the higher the ability to differentiate questions, the more participants from the high group can answer the questions correctly and the fewer test participants from the low group can answer the questions correctly. To be accepted, the value of D (discrimination: the matter of distinguishing matter) is 0.30 or more. Whereas to be declared satisfactory is 0.40 and above.

Table 2 Distinguishing Power Test Results

No	Criteria	n (%)
1	Questions received/Good (D>1,00)	9 (5,00)
2	Questions accepted (slightly revised)/Medium (0,30≤D≤0,39)	39 (21,66)
3	Revision (0,20≤D≤0,29)	97 (53,88)
4	Total Revision (D≤0,19)	35 (19,44)

Based on the results of the analysis in Table 2, it appears that the number of items that have very good differences is 9 items (5%) which means that the items can distinguish between smart test takers and less smart test takers. Furthermore, there are 21.66% have different power in both categories, which means the items can distinguish between smart test takers and less smart test takers. In addition to items that have very good different categories and good categories, some items have different abilities insufficient categories, as many as 97 items (53.88%) which mean that they must pass the revision stage. If these items have been revised, these items can be used. While the number of items that are in the category of not good as many as 35 items is equal to (19.44%) so it must be discarded (Arifin, 2017). The possibility that occurs if the quality of distinguishing power is not good is the key to the item answer is not right, the item has two or more correct answer keys, the competency measured is unclear, the deception does not function, the material is declared too difficult so many students guess, most students who understand the material being asked to think there is a misinformation in the item (Syahriandi, 2010). To increase the power of differentiation, it is better to test the

questions before use (Setyaningrum.P.M.P, et al, 2018). Some supporting factors to obtain problems with good differentiation are optimizing the process of compiling and assembling the questions following the rules, the formulation of questions must be clear so as not to cause multiple interpretations, pictures, graphs, tables, and diagrams must be clear and functional, questions using communicative language (Maenani & Oktova, 2015).

D. Difficulty level

The results of the APTIRMIKI trial test results obtained 45 items (25.00%) with a difficulty level, 87 items (48.33%) with a moderate level of question suitability and 48 items (26.66%) with an easy level (table 3). The difficulty level is a large size (Arifin, 2011). The level of difficulty of the problem is the opportunity to answer correctly at the level of need or can be agreed to discuss the problem classified as easy or difficult problems. Difficulty levels are numbers that indicate something difficult or easy to do. The magnitude of the exchange rate is between 0.00 and 1.00. The level of difficulty of items and sets of questions can be divided into three groups, namely easy, medium and difficult. To compile an exam paper that can be used to solve questions that have a balanced level of conformity, namely difficult categories of questions as much as 25%, medium categories 50% and easy categories 25% (Suharsini, 2011).

Table 3 Difficulty level

No	Kriteria	n (%)
1	Difficult (0,00- 0,30)	45 (25,00%)
2	Medium (0,31 – 0,70)	87 (48,33%)
3	Easy (>0,7)	48 (26,66%)

The difficulty level of the items is measured by the percentage of students who answer the questions correctly. If it's an easy matter, the difficulty index is higher. Questions with a p-value approaching 0 are very difficult questions, whereas questions with a p-value approaching 1 are very easy questions. An excellent index of difficulty is 0.3 to 0.7 (Akbar Rizal, 2018).

E. Distractor

Objective test in the form of multiple-choice items for each item issued in the learning achievement test has been completed with several possible answers, or often known as options or alternatives. The options or alternatives range from 3 to 5, and of the possible answers attached to each item, one of which is the correct answer (answer key), while the rest is the wrong answer. The answer to the wrong answer is commonly known as the distractor. The purpose of installing a distractor on each item is that of the many students taking the test some are interested in choosing it because they think that the distractor they have chosen is the right answer. The more students are fooled, the more distractors can carry out their functions properly. Conversely, if a distractor is installed, no one chooses, the distractor cannot perform its function properly (Solichin M, 2017).

Table 4 Distractor

No	Categori Distractor	Amount
1	Baik (0,00 – 0,25)	65 (36,11%)
2	Revisi (< 0,025)	88 (48,89%)
3	Tidak baik/ ditolak(0.000)	27 (15,00%)

The distractor of the APTIRMIKI competency tryout test functions well on



65 items (36.11%), needs to be revised by 88 items (48.89%) and rejected by 27 items (15.00%) (Table 4). Every deception can be said to function if a minimum of 5% of the total number of participants is selected (Akbar Rizal, 2018).

IV. CONCLUSIONS

1. The validity of APTIRMIKI competency test questions was 78 items (43.33%) of valid questions, while 102 items (56.67%) of questions were invalid.
2. The reliability value of the try out APTIRMIKI competency test is 0.45 so that it is included in the sufficient category.
3. APTIRMIKI competency test as many as 97 items (53.88%) passed the revision stage.
4. The results of the APTIRMIKI competency test try out obtained 87 items (48.33%) with a moderate level of difficulty.
5. The distractor of the APTIRMIKI competency tryout test functioned well on the questions totaling 65 items (36.11%) while the revision needed to be 88 items (48.89%).

V. REFERENCES

- Afiyanti, Y. (2008) 'Validitas dan reliabilitas instrumen penelitian', *Keperawatan Indonesia*.
- Akbar Rizal, M. (2018) 'Analisis Kualitas Soal di Perguruan Tinggi Berbasis Aplikasi TAP', *Jurnal Penelitian dan Evaluasi Pendidikan*, 22(1), pp. 12–23.
- Amalia, A. N., & Widayati, A. (2012). Analisis butir soal tes kendali mutu kelas XII mata pelajaran ekonomi akuntansi di Yogyakarta tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*.10(1).pp.1-26
- Arifin, Z. (2011). Evaluasi pembelajaran. Bandung: PT Remaja Rosdakarya.
- Arifin, Z. (2017) 'Kriteria Instrumen dalam suatu Penelitian', *jurnal Theorems (the original research of mathematics)*.
- Daryanto. 2010. Evaluasi Pendidikan. Jakarta: Rineka Cipta
- Hanifah, N. (2014) 'Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi', *SOSIO e-KONS*, 6(2).
- Iskandar, Akhbar;Rizal, M. (2017) 'Analisis Kualitas Soal Di Perguruan Tinggi Berbasis Aplikasi TAP', *Jurnal Penelitian dan Evaluasi Pendidikan*, 21(2).
- Istika.N, Zamsir, Rahmad Prajono. (2019). Kualitas Tes Ulangan Akhir Semester Mata Pelajaran Matematika Kelas VII Semester II SMP Negeri 12 Kendari. *Jurnal Pendidikan Penelitian Matematika*. 7(2).pp. 127-140
- Labatjo, A. I. S., Manoppo, F. P. and Berhimpon, S. L. E. (2019) 'Hubungan Nilai Ujian Try Out Lokal dengan Nilai Computer Based Test Uji Kompetensi Mahasiswa Program Profesi Dokter di Fakultas Kedokteran Universitas Sam Ratulangi Manado', *Jurnal e-Biomedik*, 7(1), pp. 49–54.
- Lukmanulhakim and Pusporini, S. L. (2018) 'Analisis Faktor yang Mempengaruhi Capaian Lulusan Uji Kompetensi Ners Mahasiswa Program Ners', *Cakrawala Pendidikan*, 37(2), pp. 306–320.
- Maenani, L., & Oktova, R. (2015). Analisis butir soal fisika ulangan umum kenaikan kelas X Madrasah Aliyah Se-Kabupaten Banjarnegara, Jawa Tengah tahun

- pelajaran 2011/2012. *Jurnal Berkala Fisika Indonesia*, 7(1), 5–11
- Makinde, O. A. *et al.* (2016) ‘Investing in health information management’, *Health Information Management Journal*, 45(2), pp. 90–96.
- Oluwatoyo, J. (2012) ‘Validity and reliability issues in educational research’, *Journal of Educational and Social Research*, 2(3).
- Permenkes, R. (2013) *Peraturan Menteri Kesehatan Republik Indonesia dan Menteri Kebudayaan Republik Indonesia Nomor 36 Tahun 2013 Nomor 1/IV/PB/2013 Tentang Uji Kompetensi Bagi Mahasiswa Perguruan Tinggi Bidang Kesehatan*. Jakarta.
- Salim, O. C. *et al.* (2016) ‘Validitas dan reliabilitas World Health Organization Quality of Life-BREF untuk mengukur kualitas hidup lanjut usia’, *Universa Medicina*.
- Setyaningrum, P.M.P, Murni Ramli, Yudi Rinanto.(2018). Analisa Kualitas Butir Soal Instrumen Assesment Diagnostic Untuk mendeteksi Miskonsepsi Siswa SMA pada Materi Virus.*Jurnal Bioedukatika*. 6(2).pp.91-101
- Solichin, M. (2017) ‘Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Soal, Intepretasi Hasil Tes dan Validitas Ramalan Dalam Evaluasi Pendidikan’, *Jurnal Manajemen & Pendidikan Islam*, 2(2), pp. 192–213.
- Srika N.P. *et al.* (2018). Analisis Butir Soal Ujian Bahasa Indonesia Buatan Guru MTsN di kabupaten Aceh Besar. *Master Bahasa*. 6(12).pp.195-202.
- Suharsini, A. (2011) *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Rineka Cipta.
- Taber, K. S. (2018) ‘The Use of Cronbach’s Alpha When Developing and Reporting Research Instruments in Science Education’, *Research in Science Education*. *Research in Science Education*, 48(6), pp. 1273–1296.
- Syahriandi. 2010. Analisis Butir Soal Ujian Semester Mata Pelajaran Bahasa Indonesia Pada SDN Seuriget Kota Langsa. *Tesis*. Banda Aceh: Program Pascasarjana Universitas Syiah Kuala.
- Zainal, A. (2012) *Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur*. Bandung. Bandung: Remaja Rosdakarya.