



Multiple-choice Test Construction Checklist for Assessing Professional Development Outcomes

Amber Halliburton
Deborah Hamby

This CASEtool describes the development and use of the *Multiple-Choice Test Construction Checklist*. A multiple-choice test is a frequently used option for assessing learning outcomes in professional development as it can be scored quickly and provides immediate feedback to the learner and instructor. Formulating sound multiple-choice questions is a time-consuming, iterative process. In this CASEtool, information is provided about the components, construction, and evaluation of multiple-choice tests. The checklist provides a set of evidence-informed criteria to ensure continuity across items and streamline the process of creating quality multiple-choice tests for assessing professional development outcomes.

INTRODUCTION

The purpose of this CASEtool is to describe the development and use of the *Multiple-Choice Test Construction Checklist* to improve the quality of test questions used to assess professional development learning outcomes. Multiple-choice tests are frequently used to assess knowledge and skills obtained from early intervention professional development. Using evidence-informed criteria to develop and evaluate test questions helps ensure a quality assessment (Wilson, 2013). The checklist that follows synthesizes and extends evidence-informed guidelines commonly used to develop multiple-choice test questions across a variety of fields (Braddom, 1997; Considine et al., 2005; Gierl et al., 2017; Haladyna, 2004; Haladyna et al., 2002; National Board of Medical Examiners (NBME) 2021; Scully 2017).

A comprehensive literature search was conducted to develop the tool. Articles were reviewed for relevancy and indicators of high-quality test questions were collected and categorized. Table 1 shows test question indicators across articles and guides. The *Multiple-Choice Test Construction Checklist* is a compilation of the most frequently occurring guidelines. The completed checklist was compared against multiple-choice item writing guidelines from Haladyna, Downing, and Rodriguez (2002). The items were organized based on typical test question development, which groups stems and answer choices together. Many fields, including early intervention and early childhood special education, rely heavily on scenario-based application of knowledge, so a section was created to include the indicators needed for quality scenario-based questions. Some guidelines related to the uniformity of the test as a whole and were grouped together in an overall section on the checklist. The analysis section provides steps in evaluating item difficulty and the

CASEtools is an electronic publication of the Center for the Advanced Study of Excellence in Early Childhood and Family Support Practices, Family, Infant and Preschool Program, J. Iverson Riddle Developmental Center, Morganton, NC. CASE is an applied research center focusing on the characteristics of evidence-based practices and methods for promoting utilization of practices informed by research.

Copyright © 2022
Center for the Advanced Study of Excellence
in Early Childhood and Family Support Practices

Table 1. Guideline characteristics by reference source.

Source	Stems				Answer Choices				
	Contain a single main idea/concept	Align with expected learner knowledge/objectives	Written in question format	Clearly and positively explain what is being asked	Ask to apply knowledge in a realistic situation	Are homogeneous in content, form, length, and number of choices	Have only one correct/best answer choice	Include plausible and relevant distractors	Have logical order and varied correct answer placement
Bothell (2001)	X		X	X	X	X	X	X	X
Braddom (1997)		X		X	X	X	X	X	
Brame (2013)	X		X	X		X	X	X	X
Butler (2018)		X		X		X	X	X	
Carriveau (2016)	X	X	X	X		X	X	X	X
Considine et al. (2005)	X	X		X		X	X	X	X
Gierl et al. (2017)						X	X	X	X
Guala (2020)	X	X	X	X		X	X	X	X
Haladyna (2004)	X	X		X	X	X	X	X	X
Malamed (2019)				X		X		X	X
NBME ^a (2021)	X		X	X	X	X	X	X	X
Rauschert et al. (2019)	X	X	X	X	X	X		X	
Ruby (2015)		X	X	X		X			X
Scully (2017)		X		X	X	X	X	X	
Towns (2014)	X	X		X		X	X	X	X
Woodford & Bancroft (2005)				X		X	X	X	X
Zimmaro (2016)	X	X		X	X	X	X	X	X

^a National Board of Medical Examiners

performance of the correct answer choice and distractors to determine their appropriateness. The resulting checklist provides a guide for efficiently developing and evaluating test items to ensure intended knowledge and skill-based outcomes are achieved.

BENEFITS OF USING MULTIPLE-CHOICE TEST QUESTIONS

Administrators and professional development providers want to understand the degree to which the intended learning is achieved from engaging in professional development experiences (Knowles, 1980). Assessment activities included in professional development help ensure the desired outcomes are achieved. Multiple-choice tests are often used because they are considered an effective, reliable, and objective way to test the participants' knowledge and comprehension (Haladyna et al., 2019; Lenchuk & Ahmed, 2021; Quagrains & Arhin, 2017; Scully, 2017; Wise, 2020). Multiple-choice tests can be easily included in digital, face-to-face, and hybrid learning models and are cost- and time-efficient in terms of administering and scoring. Although multiple-choice questions are most often used to assess gains in information and knowledge (Shakun et al., 1979), they can also assess comprehension, application, and analysis (Kheyami et al., 2018)

when presented in a real-world context. Scenario-based questions are especially beneficial for professional development participants who must apply content knowledge to real-life situations.

STRUCTURE AND FORMAT OF MULTIPLE-CHOICE TEST QUESTION

Components of Multiple-Choice Questions

While multiple-choice questions can take various formats, all include a stem, a correct answer choice, and distractors (Haladyna, 2004). The stem is the question and may range from simple recall to a detailed scenario. The answer choices refer to both the correct and incorrect answers. Distractors are often used to refer to incorrect answer choices designed to appeal to learners who may not fully grasp the presented content (Gierl et al., 2017).

Most multiple-choice questions have a single best response (NBME, 2021). Single best response multiple-choice questions have a stem and usually four or five answer options with one clear, correct response (Bothell, 2001; Braddom, 1997; Carriveau, 2016; Haladyna, 2004). Test-writers should develop questions that align with "well-defined learning objectives" for the material being taught (Towns, 2014, p. 1427) and include a single correct response (NBME, 2021).

Constructing a Quality Stem

Each question should focus on a single concept or idea (Haladyna, 2004) and avoid isolated facts (NBME, 2021). The stem presents the problem to be solved (Carriveau, 2016; Considine et al., 2005) and should be written in a question format (e.g., What is the definition of assessment?) instead of a statement (e.g., The definition of assessment is ____?) (Bothell, 2001; Carriveau, 2016). Using a question format allows the learner to focus on answering the question instead of holding the partial sentence in working memory to complete it with each option choice (Brame, 2013). Question stems should be written consistently (NBME, 2021), positively, and concisely (Brame, 2013; Carriveau, 2016; Haladyna, 2004; Zimmaro, 2016). A negative stem (e.g., Each of the following statements is true EXCEPT) can be misleading and confounding to the learner (Haladyna, 2004) and therefore not a good assessment of knowledge attainment. Similarly, extraneous or irrelevant information should not be included in the stem (NBME, 2021).

Identification Questions. Identification questions are the most common type of multiple-choice test items and assess foundational knowledge (e.g., early intervention vocabulary, characteristics of coaching, components of IDEA). These factual questions most frequently ask participants to recall a characteristic or strategy from presented content (e.g., name a characteristic of coaching) (Skakun et al., 1979). Identification items are often phrased as direct questions (e.g., Which of the following is a characteristic of relational help-giving?) and assess learning through knowledge recall and basic remembering (Bloom et al., 1956; Ruby, 2015; Skakun et al., 1979).

Scenario Questions. Scenario questions are realistic situations that allow learners to apply acquired knowledge by using a higher-order thinking process (Lenchuk & Ahmed, 2021). Engaging in a more advanced thinking process provides the opportunity to analyze, synthesize, evaluate, and engage in content-specific decision-making (Bloom et al., 1956; Ruby, 2015; Temiz, 2020). Questions that require higher-order thinking enhance learning and improve future performance (Butler, 2018). Scenario questions should provide only the information needed to answer the item since extraneous information can mislead the learner. The application of content knowledge to novel situations is especially important for early intervention practitioners as each family presents a unique opportunity to apply knowledge.

Constructing Quality Answer Choices

Answer choices should be easily understood, concisely

written, and presented in a logical order (i.e., greatest to smallest, terms, or ranges). Long answer options can shift the assessment from acquired content knowledge to assessing the learner's ability to read and comprehend so should be avoided (NBME, 2021). The correct answer choice should appear randomly amongst options (e.g., all correct answers should not be option "C") (Guala, 2020; Haladyna, 2004; Zimmaro, 2016). Trick or overtly incorrect answers should be avoided. Trick answers increase ambiguity between responses and decrease the threshold of correctness. In addition, trick answer choices can cause the learner to engage in cognitive processes that are irrelevant to the intended content and may result in users internalizing false information, potentially negating learning (Butler, 2018).

Distractors should be written plainly and with only the necessary information (Brame, 2013) and be homogeneous in grammar, content, and length (Bothell, 2001; Haladyna, 2004; Malamed, 2019; Towns, 2014). Answer options should be plausible and relevant to the content of the test and stem (Ascalon et al., 2007; Bothell, 2001; Malamed, 2019; NBME, 2021), and be placed in a logical order (Brame, 2013; Carriveau, 2016; Haladyna, 2004; Towns, 2014). Well-written distractors include commonly known errors made when learning the content (Haladyna & Rodriguez, 2013; Rauschert et al., 2019) or true statements unrelated to the stem (Gierl et al., 2017).

While distractors may be partially correct, the correct answer must be entirely true (NBME, 2021). The use of *all of the above* or *none of the above* requires additional judgment from the test taker and typically should be avoided (Butler, 2018). Requiring learners to select *all of the above* or *none of the above*, or other combination answers (e.g., "I and III" or "I, II, and III") may create ambiguity and an implied ranking of correctness among the answers, potentially exceeding the actual cognition needed to answer the question (Braddom, 1997; Brame, 2013; Carriveau, 2016; Haladyna, 2004; Malamed, 2019). Learners are required to determine which answer choices are correct, partially correct, partially incorrect, or incorrect when required to select from combination answers. The focus shifts from the learning of the material to the test-takers' ability to discern amongst answer choices (NBME, 2021). Using *all of the above* or *none of the above* changes a simple question to a complex multiple-choice question (Dibattista et al., 2014), which can lead to clueing (Butler, 2018). Both content and grammar "clues" give learners hints at the correct answer enabling "test-takers to engage in strategic guessing" and producing artificially high-performance levels and lower reliability compared to

traditional multiple-choice questions (Butler, 2018, p. 324). Overall, complex multiple-choice questions are difficult to create and are not typically better than conventional multiple-choice questions (Butler, 2018).

TEST ANALYSES

Designing a test is a recursive process and requires analyzing individual test items to determine if the correct answer and distractors perform as intended. An optimal test has items “that have high discrimination with a desired range of difficulty that maximize reliability” (Haladyna et al., 2019, p. 351). Key to item analysis is the learner recognizing the correct answer from the distractors. A quick check of this can be performed by calculating the frequency of responses for each test item on piloted test data (Gierl et al., 2017). The correct answer for any item should have the highest frequency, and the distractors should have similarly shared frequency among them. If a distractor has a low-selection frequency of less than 5%, then the distractor is very improbable to test takers and should be modified (Downing, 2009; Quaigrain & Arhin, 2017). Plotting the frequency of responses per item by overall score groups creates a visual graphic to help identify distractors that are not performing as expected (Forthmann et al., 2020; Malec & Krzeminska-Adamek, 2020). If both pre- and post-test data are present, the performance between the two can be compared. The correct answer choice is expected to be selected most frequently on a post-test as compared to the pre-test where there is an expected even spread amongst choices.

More advanced item analyses, calculated by statistical program packages, exist to understand how the test items are being answered and provide item quality indexes that help make improvements (see Downing, 2009, for more detail). The item difficulty index provides information about how hard the question is to answer and ranges from 0 (very difficult) to +100 (very easy) (Ali & Ruit, 2015; Sabri, 2013). The item discrimination index is an indicator of how well an item discriminates between the learners of higher and lower ability levels, and a value of 0.40 or greater indicates excellent item discrimination (Sabri, 2013). Other helpful statistical metrics are item means and standard deviations, point-biserial correlations, and the Kuder-Richardson 20 coefficient (KR-20), which all provide information about the test reliability. Conducting test analyses are necessary to have a quality test with items of average difficulty with functional distractors (Ali & Ruit, 2015; Quaigrain & Arhin, 2017).

USE OF THE CHECKLIST

The *Multiple-Choice Test Construction Checklist* is a tool for use in the three stages of test question construction: planning, writing, and reviewing. The tool is beneficial in the planning stage as a grounding for discussion, creating questions, and ensuring alignment with content and objectives. The checklist serves as a criteria guideline for further iterations and edits during the writing process. During the review process, this tool provides a systematic approach for assessing the performance of test questions. It is an ideal accompaniment for revising test questions especially after test item analysis.

COMPLETING THE CHECKLIST

The *Multiple-Choice Test Construction Checklist* includes guidelines for writing, revising, and evaluating multiple-choice test questions. The checklist consists of four sections describing the main components of multiple-choice assessment: (a) stem, (b) answer choices, (c) scenario-specific questions, (d) overall components, and (e) analysis. For each indicator, the user is asked to indicate whether (Yes/No) the specific criteria is present in each question. In addition, space is provided for noting any areas of concern or suggestions for improvement. When all items have been answered yes, the test should reflect well-written test question.

Individuals or groups can use the *Multiple-Choice Test Construction Checklist*. Professional development coordinators can use this checklist to align content and questions. This checklist can be used throughout the test development process, including when developing, revising, and analyzing test items. Analysis of pilot test data ensures each item is performing as expected and is a good measure of knowledge attainment.

CONCLUSION

The *Multiple-Choice Test Construction Checklist* facilitates a systematic and objective process for developing multiple-choice test questions to assess professional development outcomes. It also provides guidance for analyzing the performance of questions so that they can be revised and strengthened in a timely manner. Given the resources dedicated to the range of professional development activities typically provided in early intervention and early childhood special education, it seems particularly important to attend to how instruments are developed that measure the learning outcomes of professional develop-

ment. Using an evidence-informed checklist to develop high-quality multiple-choice questions can yield important information for facilitators, administrators, and learners.

REFERENCES

- Ali, S. H., & Ruit, K. G. (2015). The impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. *Perspectives on Medical Education, 4*(5), 244-251. <https://doi.org/10.1007/s40037-015-0212-x>
- Ascalon, M. E., Meyers, L. S., Davis, B. W., & Smits, N. (2007). Distractor similarity and item-stem structure: Effects on item difficulty. *Applied Measurement in Education, 20*(2), 153-170. <https://doi.org/10.1080/08957340701301272>
- Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956) *Taxonomy of educational objectives: Handbook I: Cognitive domain*. Longmans, Green, & Company.
- Bothell, T. W. (2001). *14 rules for writing multiple-choice questions*. Brigham Young University Faculty Center. <https://testing.byu.edu/handbooks/14%20Rules%20for%20Writing%20Multiple-Choice%20Questions.pdf>
- Braddom, CL. (1997). A brief guide to writing better test questions. *American Journal of Physical Medicine & Rehabilitation, 76*(6), 514-516. <https://doi.org/10.1097/00002060-199711000-00015>
- Brame, C. (2013). *Writing good multiple choice test questions*. Retrieved February 24, 2022 from <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>
- Butler, A. C. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning? *Journal of Applied Research in Memory and Cognition, 7*(3), 323-331. <https://doi.org/10.1016/j.jarmac.2018.07.002>
- Carriveau, R. S. (2016). *Connecting the Dots: Developing Student Learning Outcomes and Outcomes-Based Assessment*. Stylus Publishing.
- Considine, J., Botti, M., & Thomas, S. (2005). Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian, 12*(1), 19-24. [https://doi.org/10.1016/S1322-7696\(08\)60478-3](https://doi.org/10.1016/S1322-7696(08)60478-3)
- DiBattista, D., Sinnige-Egger, J. A., & Fortuna, G. (2014). The “none of the above” option in multiple-choice testing: An experimental study. *The Journal of Experimental Education, 82*(2), 168-183. <https://doi.org/10.1080/00220973.2013.795127>
- Downing, S. M. (2009). Statistics of testing. In S. M. Downing & R. Yudkowsky (Eds.), *Assessment in health professions education* (pp. 93-117). Routledge.
- Forthmann, B., Förster, N., Schütze, B., Hebbecker, K., Flessner, J., Peters, M. T., & Souvignier, E. (2020). How much g is in the distractor? Re-thinking item-analysis of multiple-choice items. *Journal of Intelligence, 8*(1), Article 11. <https://doi.org/10.3390/jintelligence8010011>
- Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research, 87*(6), 1082-1116. <https://doi.org/10.3102%2F0034654317726529>
- Guala, B. (2020, September 30). *Writing effective multiple choice questions*. UCONN ECAMPUS: Knowledge Base. <https://kb.ecampus.uconn.edu/2020/09/30/writing-effective-multiple-choice-questions-2/#>
- Haladyna, T. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Routledge.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309-333. https://doi.org/10.1207/S15324818AME1503_5
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Haladyna, T. M., Rodriguez, M. C., & Stevens, C. (2019). Are multiple-choice items too fat? *Applied Measurement in Education, 32*(4), 350-364. <https://doi.org/10.1080/08957347.2019.1660348>
- Kheyami, D., Jaradat, A., Al-Shibani, T., & Ali, F. A. (2018). Item analysis of multiple choice questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos University Medical Journal, 18*(1), e68-e74. <https://doi.org/10.18295/squmj.2018.18.01.011>
- Knowles, M. S. (1980). *The modern practice of adult education: From pedagogy to andragogy* (Revised and updated). Cambridge Adult Education Company.
- Lenchuk, I., & Ahmed, A. (2021). Tapping into Bloom taxonomy’s higher-order cognitive processes: The case for multiple choice questions as a valid assessment tool in the ESP classroom. *Arab World English Journal: Special Issue on Covid 19 Challenges April 2021*. Advance online publication. <https://ssrn.com/abstract=3851684>
- Malamed, C. (2019, September 16). *10 Rules For Writing Multiple Choice Questions*. The eLearning Coach.

- https://thelearningcoach.com/elearning_edge.
- Malec, W., & Krzeminska-Adamek, M. (2020). A practical comparison of selected methods of evaluating multiple-choice options through classical item analysis. *Practical Assessment, Research, and Evaluation*, 25(1), Article 7. <https://doi.org/10.7275/16672464>
- McKeachie W.J. (1999) Teaching, learning, and thinking about teaching and learning. In J. C. Smart & W. G. Tierney W.G. (Eds.), *Higher education: Handbook of theory and research* (Vol. 14, pp. 1-38). Springer. https://doi.org/10.1007/978-94-011-3955-7_1
- National Board of Medical Examiners (NBME). (2021). *NBME Item-writing guide: Constructing written test questions for the health sciences* (6th ed.). <https://www.nbme.org/item-writing-guide>
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1), Article 1301013. <https://doi.org/10.1080/2331186X.2017.1301013>
- Rauschert, E. S., Yang, S., & Pigg, R. M. (2019). Which of the following is true: We can write better multiple choice questions. *The Bulletin of the Ecological Society of America*, 100(1), Article e01468. <https://doi.org/10.1002/bes2.1468>
- Ruby, J. (2015, March 8). How to write great quizzes. *TD Magazine*. <https://www.td.org/magazines/td-magazine/how-to-write-great-quizzes>
- Sabri, S. (2013). Item analysis of student comprehensive test for research in teaching beginner string ensemble using model based teaching among music students in public universities. *International Journal of Education and Research*, 1(12), 1-14. <https://www.ijern.com/journal/December-2013/28.pdf>
- Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research, and Evaluation*, 22(1), Article 4. <https://doi.org/10.7275/swgt-rj52>
- Skakun, E. N., Nanson, E. M., Kling, S., & Taylor, W. C. (1979). A preliminary investigation of three types of multiple choice questions. *Medical Education*, 13(2), 91-96. <https://doi.org/10.1111/j.1365-2923.1979.tb00928.x>
- Temiz, B. K. (2020). Assessing skills of identifying variables and formulating hypotheses using scenario-based multiple-choice questions. *International Journal of Assessment Tools in Education*, 7(1), 1-17. <https://doi.org/10.21449/ijate.561895>
- Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education*, 91(9), 1426-1431. <https://doi.org/10.1021/ed500076x>
- Wilson, C. (2013). *Credible checklists and quality questionnaires: A user-centered design method*. Newnes.
- Wise, M. J. (2020). The effective use of negative stems and “all of the above” in multiple-choice tests in college courses. *Journal of Education, Teaching and Social Studies*, 2(4), 47–58. <https://doi.org/10.22158/jetss.v2n4p47>
- Woodford, K., & Bancroft, P. (2005, January). Multiple choice questions not considered harmful. In A. Young & D. Tolhurst (Eds.), *Proceedings of the 7th Australasian Conference on Computing Education* (Vol. 42, pp. 109-116). Newcastle, Australia. <https://dl.acm.org/doi/pdf/10.5555/1082424.1082438>
- Zimmaro, D. M. (2016). *Writing good multiple-choice exams*. University of Texas at Austin Faculty Innovation Center. <https://facultyinnovate.utexas.edu/sites/default/files/writing-good-multiple-choice-exams-fic-120116.pdf>

Authors

Amber Halliburton, Ed.D. and Deborah Hamby, MPH are Researchers at the Family Infant and Preschool Program, and Investigators at the Center for the Study of Excellence in Early Childhood and Family Support Practices at J. Iverson Riddle Developmental Center in Morganton, North Carolina.

Multiple-Choice Test Construction Checklist for Assessing Professional Development Outcomes

Amber Halliburton & Deborah Hamby

Name of Reviewer _____ Date _____

Checklist Descriptions

This checklist includes guidelines for writing, revising, and evaluating multiple-choice test questions. This checklist includes four subsections describing the main components of multiple-choice assessment: (a) stem, (b) answer choices, (c) scenario specific questions, (d) overall components, and (e) analysis. The *stem* is the test question, including relevant content and context. The *answer choices* are the set of options, including the most correct option. Answer choices include distractors, which are the incorrect answer options among the answer choices. *Scenario stems* include anecdotal examples to offer an application of learned material. *Overall* items pertain to the uniformity of the test as a whole. The *analysis* section contains items to help question developers evaluate the performance of the questions.

Use of the Checklists

This checklist can be used throughout the test question development process, including when first developing questions, as a guide for conducting an internal review of the questions, analyzing pilot data of the questions to ensure they are performing as quality measures of knowledge attainment, and while revising the questions. This checklist can be used by individuals or teams to develop measures of knowledge transfer occurring as a result of participating in professional development experiences, trainings, online learning modules, or courses.

To use the checklist, follow the guidance below:

- Review the checklist prior to beginning the test-writing process and familiarize yourself with the criteria for high-quality test questions.
- As you write questions, consider the learning objectives for the training and how the questions align with the training content. Test questions should address the training objectives and content. Questions that do not perform as expected could be an indication that the training does not sufficiently address the content in the questions.
- When developing questions, use each checklist item to make sure each question and the test as a whole follow the guidance for high-quality questions.
- Use the *Examples/Notes* column to indicate which test items need to be revised and in what way.
- Pilot the questions with a group of 20-30 individuals who are novice learners about the topic and conduct an analysis to ensure the questions are performing as expected and are good measures of knowledge transfer.
- Revise the questions as needed based on the analysis conducted.

Directions: Review all test questions for each criterion. Mark *Yes* or *No* to indicate if all items on the test meet expectation. Include in the Examples/Notes section notations about items needing revisions.

Question Stems	Yes	No	Examples/Notes
Align with content and objectives.			
Focus on a single concept or idea.			
Are written in a question form, asking a single complete question.			
Present the problem to be solved without extraneous or irrelevant information.			
Contain most of the wording/phrasing for the question.			
Are written positively. (Question does not ask to indicate a NOT true option).			

Answer Choices	Yes	No	Examples/Notes
Have one clear answer choice.			
Are independent of each other.			
Are homogeneous in content, detail, and length.			
Are positively phrased and in an active (not passive) voice.			
Consistently have the same answer options (4-5 per question).			
Are free from <i>all of the above</i> and <i>none of the above</i> .			
Are presented in logical/numerical order.			
Are plausible and relevant to the content.			
Have varied correct answer placement across items.			
Include distractors representative of common errors made by novice learners.			
Avoid giving clues to the correct answer.			

Scenario stems	Yes	No	Examples/Notes
Represent varied ethnicities and genders.			
Include common examples of the application of material frequently used or understood by the learner.			
Are free of extraneous or irrelevant information.			

Overall	Yes	No	Examples/Notes
All questions are independent of each other.			
No items are trick test items.			
Vocabulary matches learners expected level of understanding.			
Items have correct grammar and punctuation.			

Analysis	Yes	No	Examples/Notes
A panel of experts understood and easily answered all questions.			
Test has been piloted with a small sample of participants.			
Correct response for each question was the most frequently selected response.			
Incorrect responses were similarly spread across distractors.			
Identified items that were too easy (95% or more answered correctly) or difficult (less than 30% answered correctly) were revised.			
Distractors that were not selected at least 5% were eliminated or revised.			
Distractors chosen more frequently than the correct answer were revised.			
Overall post-test score was higher than the overall pre-test score.			

Reference: Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-333. https://doi.org/10.1207/S15324818AME1503_5