

Analysis of Validity and Consequence of Dictation in TEM4

Qian Xu

Shanghai International Studies University, Shanghai, 200083, China

Abstract

As an integrative task involving various language skills and competencies, dictation has been employed in TEM4 (Test for English Majors, Grade 4) to assess test takers' listening ability, spelling skills and writing mechanics. Therefore, it has long been an issue under discussion whether TEM4 dictation, a high-stakes test item, has measured the language skills it purports to and how it has affected English teaching and learning. In this article, the author employs the AUA (Assessment Use Argument) model to analyze the validity and consequence of TEM4 dictation, in the hope of shedding some light on the reform of TEM4.

Keywords

TEM4 dictation, AUA model, Validity, Consequence.

1. Introduction

Since the launch of Test for English Majors Grade 4 (TEM4) in 1992, the item of dictation has been employed to assess the test takers' listening ability, spelling skills and writing mechanics. As a long-standing item in one of the predominant English tests in China, dictation has been playing an important role in English teaching and learning at the tertiary level, and considerable research has been conducted on its validity, reliability and consequence.

Therefore, in this paper, I will look into the validity and consequences of dictation, analyzing its intended interpretation and use, as well as some of the unintended consequences and misinterpretations. In the analysis, Bachman and Palmer's Assessment Use Argument Model (AUA), which offers a consistent framework on an argument basis, will be employed to justify the use and interpretation of dictation, to identify possible problems and their causes, and to seek solution to the problems.

2. Dictation in TEM4

Test for English Majors Grade 4 (TEM4), a nationwide test for college sophomores majoring in English in China, has received extensive recognition from test takers, relevant institutions and society at large and is playing an increasingly important role in English language teaching and learning at the tertiary level (Jin & Fan, 2011).

As an integrative task involving various language skills and competencies, dictation is believed to be an effective method in both language teaching and testing. Being an item employed since the launch of TEM4, the dictation task aims at measuring the listening ability of test takers, as well as their ability to spell and to write in a standard format. In the course of its development in more than thirty years, the dictation part has undergone several revisions and the current version is a passage of around seventy words, read aloud for four times. The first reading is done at normal speed and the test takers are expected to listen for the general idea; for the second and third readings, the passage will be read part by part, with intervals of fifteen seconds, for the test takers to write down exactly what they have heard, including proper punctuation; the last reading is done at normal speed again, for the test takers to check through

what they have written down. Afterwards, the test takers will have one more minute to check their work.

In the marking of TEM4 dictation, the passage is divided into five chunks, with each chunk accounting for two points. Half a point will be deducted if a mistake is made, and for one chunk, the highest score deducted will be two points. The same mistake in the passage will be counted only once.

However, information provided in the Assessment report of TEM4 is very limited, just a certificate, with *Excellency* referring to a score above 80, *Good* referring to that from 70 to 80, and *Pass* equivalent to a score between 60 and 70. The test takers will receive no certificate if they obtain a score below 60. No feedback is provided on the test takers' performance on individual tasks, including the dictation part. In this sense, test takers receive no feedback on their performance on the dictation part in TEM4, which, inevitably, will lead to some detrimental effects, both on its validity and on its consequences.

3. Assessment Use Argument

Assessment Use Argument (AUA) provides an overarching framework to link inferences from test takers' assessment performance to the assessment records, interpretations, decisions, and consequences, which guides not only the design and development of language assessments, but also the interpretation and use of language assessment results. These five interrelated links can be either accounted in a top-down approach from consequences to assessment performance for test development, or in a bottom-up manner for the interpretation and use of language assessment results. (Bachman & Palmer, 2010) Therefore, with the AUA Model, we can seek warrant and rebuttal to back and rebut the claims made in the four parts—consequences, decisions, interpretations, and assessment records, to justify the use of dictation, to seek problems, and to find solutions.

First of all, we shall consider the consequences of TEM4 dictation, to find out what the consequences are, to identify possible negative impact and its causes, and to work out approaches to promote beneficial washback. Next, we will look into the decisions to be made based on a test, to find out whether they are value-sensitive; that is, whether they comply with relevant educational and societal values, as well as laws, rules and regulations. Moreover, fairness of the decision to all test takers is also essential to a valid test. Thirdly, when interpreting the ability to be assessed, we have to find out whether the score interpretations are meaningful with respect to the learning syllabus or some language ability theories, impartial to all test takers, generalizable to target language use domains, relevant to the decisions to be made and sufficient to make certain decisions. Last but not least, assessment records shall be consistent, regardless of different aspects of assessment procedure or different groups of test takers.

4. Validity and Consequences of Dictation in TEM4

Validity is the extent to which a test measures what it is meant to measure, and for dictation in TEM4, it is supposed to measure test takers' listening ability, spelling skills and writing mechanics. Consequences refer to the effects of the assessment on different stakeholder groups, such as teachers and students. To illustrate, whether preparing for the dictation task in TEM4 will facilitate or hinder classroom teaching and English learning.

In the following part, validity and consequences of dictation in TEM4 will be justified by addressing the four parts in the AUA Model, from consequences to assessment records.

4.1. Consequences

The purpose of TEM4 dictation is to assess the listening ability, spelling skills, and writing mechanics of college English majors in the foundation stage (first and second year in university), to examine whether they have met the required levels specified in the National College English Teaching Syllabus. Meanwhile, TEM4 may facilitate the evaluation of English teaching, to evaluate how the teaching syllabus is implemented in different universities, and to help promote the reform on English teaching methodology and curriculum design at the tertiary level.

In this sense, the consequences of TEM4 dictation are beneficial to the students (test takers), teachers and schools in two aspects. On one hand, TEM4, as a high-stake test, may motivate students to work harder, guide students in their learning and provide them with feedback to help identify their strengths and weaknesses. As for the teachers, TEM4 may help improve their teaching, providing both guidance and feedback on the teaching outcome. Moreover, with feedback from TEM4, schools can improve their English teaching programs accordingly. On the other hand, dictation is believed to be an effective method employed in both teaching and testing, and it is an integrative task involving various language skills and competencies, such as auditory discrimination, auditory memory span, lexical recognition, spelling, grammar knowledge and discourse comprehension (Heaton, 1988). Therefore, TEM4 dictation may lead to some positive consequences in that the test preparation activity may help improve English teaching effect and students' language ability. In this sense, TEM4 dictation has achieved the beneficial consequences.

However, some practices of TEM4 dictation may have negative consequences. First, the feedback provided in the assessment reports of TEM4 is very limited, only a certificate with a certain grade, which is not clear enough for test takers to find out their strengths and weaknesses in English learning, or for teachers to evaluate their teaching outcome. In this way, the positive consequences, such as motivating students to overcome their weaknesses, can hardly be achieved due to the limited information provided in TEM4 test reports. Moreover, due to the lack of communication between test designers and test takers, some students may have no idea about the purpose of TEM4 dictation. Unlike other tasks such as reading comprehension, dictation is an integrative task measuring the comprehensive language ability of test takers, so it is natural that some test takers may doubt the effectiveness of dictation, claiming that it measures nothing but making them anxious, for the dictation task is very tense in that test takers are required to be highly concentrated when listening to the dictation passage.

4.2. Decisions

As a section in TEM4, dictation accounts for 10% of the overall score, but no feedback will be provided on this section, and the decisions to be made are all based on TEM4 score, rather than test takers' performance on the dictation task only. In this sense, here we will, instead, look into the decisions made on the interpretation of TEM4 assessment report.

The chief decision to be made on TEM4 is certification—to certify students' level of language ability at the end of the foundation stage in their college education program. As a national test, TEM4 is designed by experts in English, most of whom are teachers and administrators in the participating universities, so it can be ensured that the existing educational and societal values and relevant legal requirements in the test development will be taken into consideration. Moreover, the test syllabus has been published and sample tests are provided, so the test takers should be fully informed about how the decisions will be made. Last but not least, the test takers will be classified only according to the cut score and decision rules, and nothing else will be taken into consideration. In this sense, the decision made should be equitable to all test takers. However, in some cases, TEM4 test score may be used to make other decisions, such as whether the test taker can be awarded the Bachelor's Degree in English Language and Literature,

whether the test taker is eligible for a certain job, and whether the test taker can apply for a postgraduate program. These decisions can hardly be justified, since TEM4 is only a test measuring test takers' language ability at the end of the sophomore year. It can hardly predict test takers' performance upon graduation, not to mention their ability in future careers. Therefore, test developers and users of TEM4 should be very cautious when making other decisions based on TEM4 test score.

4.3. Interpretations

When interpreting the language ability assessed by TEM4 dictation, we will consider its meaningfulness, impartiality, generalizability, relevance and sufficiency. To begin with, the construct for TEM4 dictation is clearly stated in the test syllabus of TEM4, which is to measure test takers' listening ability, spelling skills, and writing mechanics. When taking TEM4 dictation, test takers are expected to write down exactly what they have heard, including proper punctuation; and in the marking of TEM4 dictation, mistakes of various kinds will lead to the loss of marks, such as missing words, grammar mistakes, and spelling mistakes. In this regard, TEM4 dictation measures not only listening ability, but also spelling skills and writing mechanics, just as stated in the test syllabus. Second, the interpretation of TEM4 dictation score is impartial in that the test content and format are carefully designed to avoid bias against any group of test takers. Since the test takers are university students, their target language use domains are either language teaching and learning domains or real-life tasks. In this sense, the theme of TEM4 dictation passage is always a familiar topic, such as the introduction to a common subject (Gestures in 2022 TEM4) and the definition of a common term (Chinese civilization in 2023 TEM4).

However, in terms of generalizability, relevance and sufficiency, there are some factors that may affect the interpretations of the ability assessed by TEM4 dictation.

First, the gap between TEM4 dictation task and real-life task may affect the generalizability of the test. The interpretation will be more generalizable if the characteristics of the assessment task correspond closely to that in the target domain. However, the dictation task requires test takers to listen to a passage for four times to note down the exact words in it, which seems to be quite different from a real-life situation. On one hand, in real life, like a seminar, it is quite unlikely that a definition will be repeated four times; on the other hand, when taking notes in the classroom, students are expected to write the general idea rather than the exact words. In this sense, the difference between dictation and real-life tasks may affect its generalizability. Second, the relevance and sufficiency of interpretation can hardly be ensured if some unauthorized decisions are to be made on TEM4 score, such as whether the test takers are qualified for some jobs. Since TEM4 is a highly recognized English test in China, it is quite likely that it may be used for other purposes, such as selection and placement, both for students and for teachers. However, these are not the inferences that can be drawn from the test results, and will affect the relevance and sufficiency of interpretation as a result.

4.4. Assessment records

Tremendous efforts have been made to ensure the consistency of TEM4 assessment records, such as to follow consistent administrative procedures, to carry out strict rater training and to employ double rating. In the marking of TEM4 dictation, all raters will gather together to receive meticulous training and to rate the dictation together. In this way, TEM4 dictation score should be consistent across different test procedures and across different group of test takers.

5. Suggestions on the Reform of TEM4 Dictation

From the above analysis, we may arrive at the conclusion that TEM4 dictation is of fairly high validity and its consequences on teaching and learning are mainly positive. However, to further improve its validity and consequences, reform can be conducted in the following aspects.

To begin with, a detailed assessment report is of great importance. The feedback information provided by TEM4 is quite limited, just a grade, and no feedback is provided on the dictation section. This will definitely affect the validity and consequences of TEM4 dictation, since teachers and students know little about the students' performance in each part, and as a result they can hardly adjust their teaching and learning plan accordingly.

Next, communication should be promoted between test designers and test stakeholders, such as students, teachers and even school administrators. By way of publicizing the test syllabus, sample tests and marking criteria, stakeholders of various fields can better understand the test purpose, test score interpretation, and relevant decisions. In this way, more efficient teaching plans can be developed, test takers can learn more effectively, and test score users like university authorities and companies can make more reasonable decisions.

Thirdly, standard use of test report should be promoted and decisions made should be justified. TEM4 is an assessment of test takers' English proficiency at the end of the foundation stage (sophomore year) in the university, and TEM4 dictation aims at measuring test takers' listening ability, spelling skills and writing mechanics. In this sense, decisions made on TEM4 dictation performance should be reasonable, based on the relevant ability assessed by TEM4 dictation.

Last but not least, the test method of dictation can be improved to simulate real-life tasks. The current form of dictation is to listen to a passage for four times and to write down every word of it, which is quite different from the real-life task of taking notes. Therefore, the test format of dictation can be revised to simulate real-life situations, such as to listen only once and to summarize rather than to write down the exact words heard.

6. Conclusion

As a long-standing item in TEM4, dictation has played an important role in measuring test takers' language proficiency and exerting positive consequences on English teaching and learning. However, it has also attracted some criticisms and challenges. With the AUA Model, we can better prove and justify the validity and consequences of TEM4 dictation, making suggestions on the reform and improvement of TEM4 dictation.

Acknowledgments

This article is funded by the Fundamental Research Funds for the Central Universities in China (Research Grant No. KY01X0222017082).

References

- [1] Bachman, L. & Palmer, A. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.
- [2] Jin, Y. & Fan, J. (2011). Test for English Majors (TEM) in China. *Language Testing* 28(4): 589-596.
- [3] NACFLT (2004a). *Syllabus for TEM4*. Shanghai: Shanghai Foreign Language Education Press.
- [4] NACFLT (2000). *Syllabus for University English Language Teaching*. Shanghai: Shanghai Foreign Language Education Press. Beijing: Foreign Language Teaching and Research Press.