Appendix A: Example WARP V Assessment Review

Confusing Words from Dave's ESL Cafe

Dave's ESL Café is a free website where ESL teachers can find a variety of resources from job advertisements to quizzes. In this quiz on "confusing words," you can find 10 multiple-choice questions that test students' ability to differentiate the meanings of commonly confused words; in most cases as a result of similar spelling or sounds. Students read through the short 10 questions and fill in the blanks in the sentences, choosing one of the two options. After hitting the submit button, the website produces the results.

Example:

1. Dave had chocolate ice cream for _____.
    a) desert
    b) dessert

Benchmark	Criterion	Assessment	Improvement
Washback	Are students learning from this test by taking it or preparing for it? Does the test have any potential detrimental effects?	There is limited opportunity for washback in terms of preparation. There is limited opportunity to learn from the test. After submitting the answers, the review of correct and wrong answers on the website is minimal (i.e., simple check or cross marks). The review simply produces a percentage of correct answers and indicates which ones are correct or wrong.	If teachers can prepare students on this content and then students are tested, it may achieve a preparation effect. The test could explain, for example, why many people may choose a particular incorrect answer, and where the misunderstanding may stem from. For the example Question #1, an explanation could be added: Dessert, which we eat, has two "s" letters in it and is pronounced this way.

Authenticity	Is the language in the test as natural as possible? Are items contextualized rather than isolated? Are the topics meaningful (relevant, interesting) for learners? Do the tasks approximate real-world tasks?	The quiz has limited authenticity because the items are isolated in meaning, and do not constitute a whole together. There is also no common theme across the questions and they may not be interesting to students. The tasks do not necessarily approximate real-life tasks.	Questions could be written to build on each other. The pronunciation of each word or some accompanying pictures could also be provided so that students can use them as contextual cues.
Reliability	Would the students score the same if they took the test again? Are there any outside factors that might impede students' performance?	Multiple-choice tests are generally strong in reliability, and this test is no exception. It would not be affected by a rater bias, and the test would yield reliable results. Presenting only two options increases the possibility of finding the correct answer by chance.	When possible, present more than only two answers.
Practicality	Is the test affordable? Does it stay within appropriate time constraints? Is it easy to administer? Does it have a scoring/evaluation procedure that is specific and efficient?	The test is practical. It doesn't take a lot of time, and the website produces the results. It is free and easily accessible on the internet.	N/A

Validity	Is the test measuring what it is intended for or something else? Is the test measuring the intended language skill or construct?	The validity aspect might be the greatest weakness of this test. At first look, it seems to be measuring students' ability to differentiate commonly confused words. However, the words chosen do not sample from a meaningful construct and may seem random (i.e., *dessert*/*desert*, *lose*/*loose*, *capital*/*capitol*, *hear*/*here*). It is not clear why these words were selected.	To improve validity, teachers can collect commonly confused words in their own classes based on their observations and conduct whole class administration using kahoot or individual white boards.

In conclusion, the "Commonly Confused Words" quiz on Dave's ESL Café did not do so well across the WARP V benchmarks. First, the test does not teach students why certain answers are correct, missing the mark for Washback. Second, the items are not interesting or contextualized to make up a meaningful assessment, approximating real-life tasks; therefore are not Authentic. Third, it may produce reliable results because there is only one correct answer, but the possibility of finding that answer is 50%, so it is quite likely to be found by chance. It is still somewhat reliable because there is no rater bias involved. Fourth, it doesn't take a long time to prepare or take the test, so it is Practical. Finally, the test is meant to measure if students can differentiate between two commonly confused words. It measures this ability for the selected words, but the sampling of words seems random, and their relevance to students is questionable. Therefore, it suffers in Validity.

To improve this test in general, teachers can prepare their own quizzes using words that are commonly confused by their own students. They can give multiple-choice questions that would be meaningful to their students, and use the target words on an interactive game like kahoot, where students can join in on their mobile devices. Alternatively, students can have small white boards where they write their answers individually when each question is presented. These adjustments would help teachers practice and assess students' knowledge on commonly confused words in a manner that addresses all WARP V benchmarks for this activity.