

Report for Computer Simulations LLC (CSW) 2010 Post-Test Analysis

Computer administered dental licensing testing developed by Central Regional Dental Testing Service (CRDTS), Southern Regional Testing Service (SRTA) and Western Regional Examining Board (WREB).

Del Hammond, M.A., Testing Specialist

June 2011

Background for statistics used in presentation of 2010 test results

Licensing tests are used to identify individuals who meet the requirements for performing an activity or profession. There is no need to identify very high or very low performing individuals. These tests only need to separate the qualified individuals from the unqualified. The test items only need to be those that evaluate the minimum abilities required for the license. There is no need to find out who has knowledge or skill that is well above what is required for licensure. Consequently, there is no need for items that are very difficult for qualified applicants. Most of the candidates are well qualified, so it is expected that most of them should be able to correctly answer most of the questions.

Differences between general ability testing and licensing tests, regarding test question difficulty and differences in test taker knowledge and abilities, cause the formulas for reliability estimates to be of marginal value for evaluating licensing tests. The internal consistency reliability estimation formulas were designed for and are useful for educational and general ability testing. The basis for the formulas is the expected variability among candidate scores and the expected variability in the correctness of individual candidate responses. Those variabilities are very low for dental licensing testing. As a result, the formulas and “rules” for educational testing are not very useful for licensing tests. The formulas provide reliability estimates that underestimate (are lower than) true reliability.

CSW Post-Test Analysis

The CSW Management Committee held separate meetings on January 7th and February 4th of 2011 to discuss possible alternative methods for analyzing and reporting CSW test results. A different, independent testing specialist attended each meeting. Both testing specialists advised against reporting alpha estimates. Unless a test is unidimensional, alpha estimates are inappropriate and should not be reported. Analysis of CSW exams reveals that they are multidimensional. Factor analysis didn't reveal any groupings that would allow the use of stratified alpha. CSW now uses conditional standard errors of measurement and measures of decision consistency, rather than alpha estimates, to report the quality of CSW tests. These values are used for two reasons. First, because, that for the CSW test questions (items), and for well prepared candidates, alpha reliability estimates underestimate true reliability. Second, because the *Standards for Educational and Psychological Testing* (the 1999 Standards) recommend the use of conditional standard errors of measurement and measures of decision consistency. Because the goal of licensure testing is to identify qualified candidates, these measures are of specific importance and are appropriate measures of test quality.

CSW has used measures of decision consistency in the past, but this is the first use of conditional standard errors of measurement (CSEM) in reporting results. The relevant standards from the 1999 Standards are:

Standard 2.14

“Conditional standard errors of measurement should be reported at several score levels if constancy cannot be assumed. Where cuts cores are specified for

selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score”.

Standard 2.15

“When a test or combination of measures is used to make categorical decisions, estimates should be provided of the percentage of examinees who would be classified in the same way on two applications of the procedure, using the same form or alternate forms of the instrument”.

Standard 14.15

“Estimates of the reliability of test-based credentialing decisions should be provided”.

Dr. Larry R. Nelson from Curtin University of Technology, Western Australia points out, in *Some Issues Related To The Use Of Cut Scores*, that “We should compute the standard error of measurement at the cut score, employ an estimate of the percentage of examinees who have been consistently classified on both sides of the cut, and, when appropriate, use a measure of the reliability of test-based credentialing decisions”. He also states “We know that tests with a majority of easy items will often return low alpha and item discrimination values when put into action. This does not mean that the tests are necessarily deficient. As Berk (2000) pointed out, it may be partly a matter of using the wrong statistics – if our interest is with assessing mastery, as in licensing and certification efforts, then we should deploy the methods recommended in the *Standards*”(Nelson, 2007).

CSW now uses CSEMs and Peng-Subkoviak estimates (p_0) of decision consistency. These values can be found using the Lertap program developed by Dr.

Larry Nelson. It is available from Assessment Systems Corporation which also markets numerous testing publications and the Iteman item analysis program that CSW has used for many years. In his paper on the use of cut scores, Dr Nelson suggests that, for “serious decisions,” a desired value of p_0 is .85, which is the same value that CSW has previously referenced when using measures of decision consistency

Post-Test Analysis Results for CSW 2010 TestingPeriodontal

	FORM 1	FORM 2
Number of Candidates	1262	1307
Passing Percent	96.3 %	96.0 %
ρ_0 , Decision Consistency	0.966	0.967
CSEM	2.52	2.83
Cut Score	20	24
Median Score	25	30
Standard Deviation	2.67	3.28
Skew	- 0.62	- 1.21
Kurtosis	0.73	5.64

Prosthodontics

	FORM 1	FORM 2	FORM 3
Number of Candidates	878	846	845
Passing Percent	97.6 %	94.4 %	92.6 %
ρ_0 , Decision Consistency	0.916	0.956	0.926
CSEM	3.21	3.0	2.96
Cut Score	34	35	35
Median Score	43	41	41
Standard Deviation	3.68	3.56	3.6
Skew	- 0.91	- 1.17	- 0.95
Kurtosis	1.36	3.47	1.9

Note that the decision consistency values for all Prosthodontics and Periodontal forms are well above the 0.85 recommended minimum for licensure testing.

References

- American Educational Research Association, American Psychological Association.
National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Nelson, L. R. (2007). Some issues related to the use of cut scores. *Thai Journal of Educational Research and Measurement*, 5(1) (ISSN 1685-6740), 1-16.