OFFICE OF EDUCATIONAL ASSESSMENT
Improving Learning Through Assessment

OEA Report 16-03

# Chemistry 142 Placement Exam Development

*Debbie McGhee and Debbie Wiegand*
*June 2016*

## INTRODUCTION

This report describes development of the University of Washington (UW) Chemistry 142 placement exam. The exam was developed by the UW Department of Chemistry in collaboration with the Office of Educational Assessment and first administered in June 2016.

The Chemistry 142 placement exam was developed over the course of four academic quarters. Our goal was to create two parallel forms of the exam, one for initial administration and a second for retesting. Each exam would consist of approximately 25 items for a one-hour administration period.

Development occurred in two phases: item writing and forms development, and forms validation. The next section of the report provides a brief description of the methods for both phases, followed by a more detailed description of the analyses and results relating to instrument validation.

## METHOD

### Item writing and forms development

Test development began in Spring 2015 with a review of Chemistry 142 course content and identification of subject areas in which student preparation would affect likelihood of success in the course. We created a table of specifications detailing the total number of items needed by content area for two parallel 26-item test versions. An initial item pool of four-alternative multiple-choice items was created by Chemistry faculty members and we subsequently reduced this pool to 120 items using a process of review, discussion, and revision over the course of Summer 2015.

In Autumn 2015, we carried out preliminary item testing by creating six 20-item tests from the pool of 120 items, and administering the tests to 326 students enrolled in Chemistry 142. We administered the tests during a single lecture period in the first week of the quarter, allowing students approximately 35 minutes to complete the exam. Based on the results of standard item analysis, we reduced the number of item alternatives from four to three, and created three parallel test versions (A, B, and C) of 26 items each.

## Forms pilot and validation

During Winter quarter 2016, we administered the three parallel test versions (A, B, and C) to students participating in the first lab meetings of all three sections (A, B, and C) of Chemistry 142. The length of the lab allowed students to take up to 60 minutes to complete the exam. There were 40 lab meetings, and the maximum number of students per lab meeting was 24; all students within a lab were administered the same test version.

A total of 839 students completed the exam, but 22 students performed below chance (i.e., total score less than 9) and another three students dropped the course. These students were removed from the student sample, leaving $n = 814$ cases available for analysis. The number of students who completed each test version within each class section is shown in Table 1 and the specific data analyses carried out are described in the next section.

**Table 1. Number of cases by test version and course section**

| Test | Course Section | | | |
|---|---|---|---|---|
| Version | A | B | C | Total |
| A | 95 | 86 | 91 | 272 |
| B | 117 | 95 | 62 | 274 |
| C | 114 | 88 | 66 | 268 |
| Total | 326 | 269 | 219 | 814 |

# FORMS PILOT AND VALIDATION RESULTS

We carried out a variety of analyses of both item and test characteristics as described below. Analyses included examination of test difficulty, internal consistency, item quality, relationship of test scores to achievement indicators, and determination of a placement cut score.

## Test difficulty

The overall difficulty of the three tests was moderately high. The average score over all three test versions, was $Mn = 17.1$[1] out of 26 possible (66% correct), as shown in Table 2.

Students found test version C to be more difficult than versions A and B[2]. This effect was found for Sections A and B but not Section C (see Figure 1), possibly reflecting somewhat different enrollment patterns across class sections.

[1] 95% CI: 16.8, 17.4

[2] $F(2,805) = 3.43$, $p = .03$

**Table 2. Average score by test version**

| Test Version | Student performed below chance: No Mean | SD | n | SE | Yes Mean | SD | n | SE |
|---|---|---|---|---|---|---|---|---|
| A | 17.5 | 4.04 | 272 | 0.25 | 7.0 | 1.53 | 7 | 0.58 |
| B | 17.3 | 3.99 | 274 | 0.24 | | | 0 | |
| C | 16.5 | 4.08 | 268 | 0.25 | 6.6 | 1.35 | 15 | 0.35 |
| TOTAL | 17.1 | 4.05 | 814 | 0.14 | 6.7 | 1.39 | 22 | 0.30 |

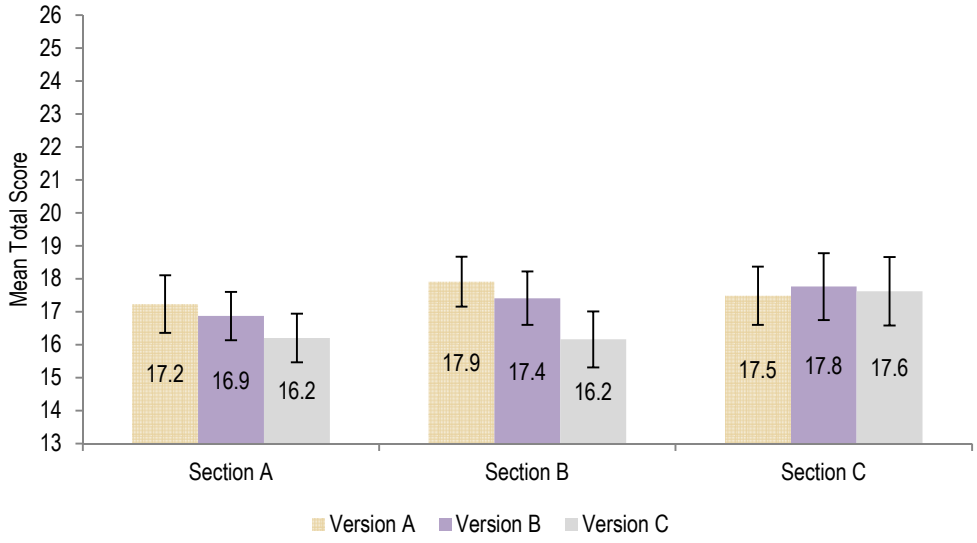


**Figure 1. Mean placement test score by test version and course section**

## Internal consistency

We examined the internal consistency of each of the three test versions by computing Cronbach's *alpha.* The resulting coefficients were .73, .71, and .71, for Versions A, B, and C, respectively. Because the exam was designed to test knowledge of up to eleven separate yet inter-related concepts, alpha coefficients of this magnitude (i.e., $\alpha < .80$) were to be expected.

## Item quality

We examined three indicators as evidence of individual item quality. Item difficulty, item discrimination, and the number of "good" alternatives are reported in Table 3 for all items.

**Item difficulty** is the proportion of all test-takers who answered the question correctly. Optimal item difficulty is generally accepted as approximately halfway between chance and 1.0; thus for a three-alternative item, optimal difficulty is $D = .67$. Difficulty levels between .50 and .69 are shaded in Table 3.

**Table 3. Item difficulties, discrimination, and number of good alternatives by test version**

| Topic | Item No. | Version A Difficulty | Version A Discrim-ination | Version A Number of good alts. | Version B Difficulty | Version B Discrim-ination | Version B Number of good alts. | Version C Difficulty | Version C Discrim-ination | Version C Number of good alts. |
|---|---|---|---|---|---|---|---|---|---|---|
| **"math"/problem solving** | | | | | | | | | | |
| | 01 | .81 | .15 | 2 | .86 | .11 | 2 | **.60** | .22 | **3** |
| | 12 | **.50** | .13 | **3** | .76 | .09 | 2 | **.63** | **.37** | **3** |
| | 21 | .49 | **.34** | 2 | **.64** | .23 | **3** | **.51** | **.40** | **3** |
| **balancing equations** | | | | | | | | | | |
| | 02 | .76 | **.39** | **3** | .77 | **.30** | **3** | .72 | .12 | 2 |
| | 13 | .85 | **.39** | 2 | .78 | **.28** | 2 | **.69** | **.37** | **3** |
| | 22 | .87 | .20 | **3** | .85 | **.26** | **3** | .83 | **.33** | **3** |
| **basic stoichiometry** | | | | | | | | | | |
| | 03 | **.66** | **.38** | **3** | .46 | **.28** | **3** | .31 | **.40** | 2 |
| | 14 | .76 | **.35** | **3** | .45 | **.43** | **3** | .46 | **.25** | **3** |
| | 23 | .23 | **.28** | 2 | .42 | **.39** | **3** | .49 | **.40** | **3** |
| | 26 | **.60** | **.38** | **3** | **.59** | **.34** | **3** | .41 | **.46** | **3** |
| **periodicity of elements** | | | | | | | | | | |
| | 04 | .76 | **.35** | **3** | .80 | **.38** | **3** | .46 | **.33** | **3** |
| **basic atomic structure** | | | | | | | | | | |
| | 05 | .83 | **.34** | **3** | .87 | .00 | 0 | .88 | .11 | 2 |
| | 15 | .87 | .17 | 2 | .96 | .16 | 1 | .72 | **.31** | **3** |
| | 24 | .71 | **.41** | 2 | **.57** | **.26** | **3** | | | |
| **covalent vs ionic bonding** | | | | | | | | | | |
| | 06 | .83 | **.28** | 2 | .73 | .13 | 2 | .36 | **.25** | **3** |
| | 16 | **.67** | .13 | 2 | .43 | .22 | **3** | .48 | **.30** | **3** |
| | 24 | | | | | | | .89 | .06 | 2 |
| | 25 | .71 | **.36** | **3** | **.63** | .23 | **3** | .80 | .11 | 2 |
| **nomenclature of inorganics** | | | | | | | | | | |
| | 07 | .76 | **.42** | **3** | .81 | .21 | **3** | .72 | **.32** | 2 |
| | 17 | .72 | .18 | **3** | .72 | **.27** | 2 | .78 | **.32** | **3** |
| **atomic mass** | | | | | | | | | | |
| | 08 | **.56** | **.25** | **3** | **.53** | **.28** | **3** | .78 | **.30** | **3** |
| **mole concept** | | | | | | | | | | |
| | 09 | .27 | **.32** | 2 | .30 | .22 | **3** | **.54** | **.36** | 2 |
| | 18 | **.53** | **.26** | **3** | .49 | **.38** | **3** | .48 | **.37** | 2 |
| **molar mass** | | | | | | | | | | |
| | 10 | .73 | **.28** | **3** | .71 | .14 | **3** | .34 | .22 | 2 |
| | 19 | .83 | **.31** | **3** | **.58** | **.29** | **3** | .84 | **.26** | 2 |
| **composition of compounds** | | | | | | | | | | |
| | 11 | **.57** | .17 | 2 | .88 | .20 | **3** | **.69** | **.40** | **3** |
| | 20 | .38 | **.26** | 2 | **.65** | .21 | **3** | **.59** | **.27** | 2 |

**Item discrimination** is the corrected item-total correlation and should equal or exceed $r_{pb}$ = .30. The average item discrimination index (corrected item-total correlation) for all items was .27 (range: .00 - .46). Discrimination indices above .245 are shaded in Table 3.

Table 3 also shows the **number of "good" alternatives** for each item, where a "good" alternative is defined as an answer choice that attracts at least 5% of the test-takers and has an alternative-total correlation of $|r| \geq .095$. The average number of "good" alternatives per item was 2.6 (range: 0-3). All items on each version had at least two "good" alternatives; instances in which there were three "good" alternatives are shaded in Table 3.

The number of individual test items meeting acceptable criteria for each of the three indices is shown in Table 4.

**Table 4. Summary of discrimination and distractor performance by test version**

| Test Version | Number of items with… | | |
|---|---|---|---|
| | Difficulty .50 to .70 | Discrimination ($r_{pb}$) ≥ .245 | Three strong alternatives |
| A | 7 | 19 | 15 |
| B | 7 | 13 | 19 |
| C | 7 | 20 | 15 |

## Relationship of test scores to performance in the course

The relationship between student test performance and subsequent course grade is an indicator of the potential effectiveness of the test in placing students into the course. Table 5 shows the average Chemistry 142 course grades and the correlations between test scores and subsequent course grades. Overall, test scores were moderately correlated with course grades ($r$ = .48), indicating that the observed test scores were good predictors of ultimate performance in the course.

**Table 5. Course grades and test score-course grade correlations by test version and course section**

| Test Version | | Course Section | | | |
|---|---|---|---|---|---|
| | | Section A | Section B | Section C | Total |
| **A** | *Mn* | 2.82 | 2.88 | 2.71 | 2.80 |
| | *SD* | 0.73 | 0.70 | 0.81 | 0.75 |
| | *n* | 95 | 83 | 88 | 266 |
| | *SE* | 0.07 | 0.08 | 0.09 | 0.05 |
| | Test Score-Course Grade *r* | .38 | .42 | .47 | .42 |
| **B** | *Mn* | 2.82 | 2.90 | 2.66 | 2.81 |
| | *SD* | 0.67 | 0.75 | 0.89 | 0.75 |
| | *n* | 116 | 91 | 60 | 267 |
| | *SE* | 0.06 | 0.08 | 0.11 | 0.05 |
| | Test Score-Course Grade *r* | .56 | .55 | .53 | .53 |
| **C** | *Mn* | 2.78 | 2.96 | 2.75 | 2.83 |
| | *SD* | 0.66 | 0.73 | 0.87 | 0.74 |
| | *n* | 110 | 85 | 62 | 257 |
| | *SE* | 0.06 | 0.08 | 0.11 | 0.05 |
| | Test Score-Course Grade *r* | .47 | .50 | .67 | .51 |

Figure 2 provides a visual representation of the overall relationship between test scores and course grades. The four left-most box-plots represent the distribution of test scores by course grade quartiles. Test scores increase in a consistent linear fashion in keeping with grade quartiles: the median test score among students in the lowest grade quartile was 14 (*Mn* = 14.0), while the median among those in the highest quartile was 20 (*Mn* = 19.8).

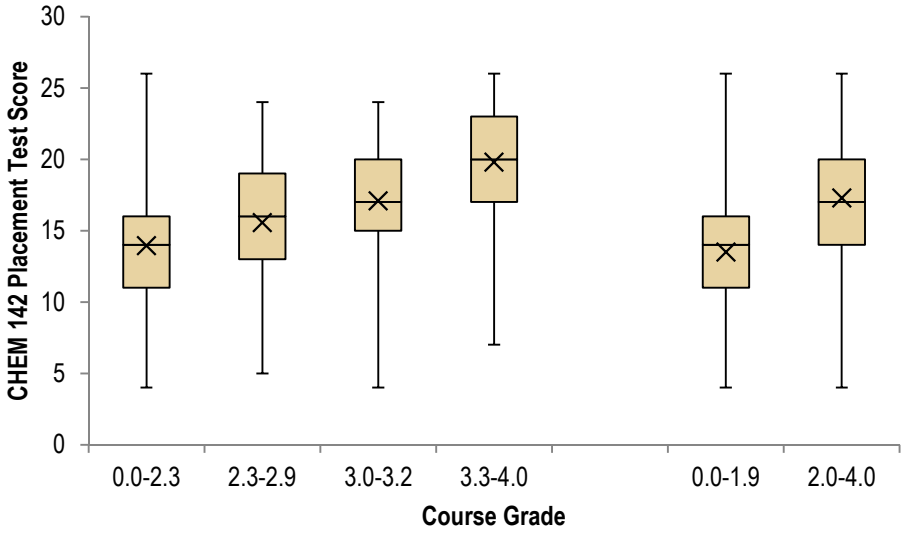


**Figure 2. Distributions of placement test scores by course grade. The leftmost boxplots display the observed grade quartiles.**

The two right-most box-plots in Figure 2 display the distribution of test scores by course pass/fail grading (i.e., grades above and below 2.0). Test scores again are consistent with course grades – among students who earned a grade below 2.0, the median test score was 14 (*Mn* = 13.5) versus a median score of 17 (*Mn* = 17.3) among those with grades equal to or greater than 2.0.

The comparison of test score to pass/fail course grades was consistent across test versions, as shown in Figure 3. The three left-most box plots in this figure represent test scores of students who earned a course grade below 2.0, while the three right-most plots represent test scores of students with grades equal to or greater than 2.0. As noted earlier, Version C proved to be slightly more difficult than Versions A and B.

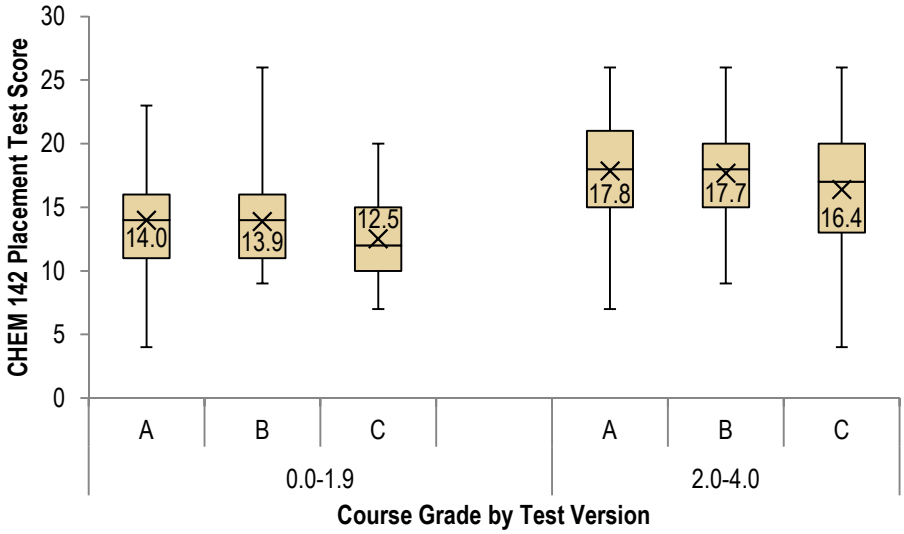


**Figure 3. Distributions of placement test scores by course grade (below or above 2.0) and test version.**

## Relationship of test scores to other achievement indicators

In addition to Chemistry 142 course grades, we examined the relationship of placement test score with other indicators of student achievement. These indicators, shown in Table 6, were total UW GPA, high school GPA, ACT, SAT, and MPT test scores, and AP scores. Particularly noteworthy were the moderately strong correlations between placement test score and ACT Science & Reasoning (*r* = .47) and AP Chemistry (*r* = .52).

**Table 6. Correlations between chemistry placement test scores and related achievement indicators by test version**

| Achievement Measure | | Test Version A | Test Version B | Test Version C |
|---|---|---|---|---|
| **Total UW GPA** | *r* | .12* | .19** | 0.11 |
| | *n* | 260 | 260 | 267 |
| **HS GPA** | *r* | 0.11 | .14* | .17** |
| | *n* | 249 | 256 | 257 |
| **ACT** | *r* | .42** | .53** | .44** |
| | *n* | 96 | 92 | 96 |
| **ACT Math** | *r* | .33** | .48** | .44** |
| | *n* | 96 | 92 | 96 |
| **ACT Science** | *r* | .38** | .58** | .47** |
| | *n* | 96 | 92 | 96 |
| **SAT Math** | *r* | .46** | .49** | .42** |
| | *n* | 214 | 221 | 206 |
| **MPT Advanced** | *r* | .41** | .42** | .28** |
| | *n* | 106 | 79 | 107 |
| **MPT General** | *r* | .28 | .42* | .42** |
| | *n* | 26 | 35 | 37 |
| **AP Chemistry** | *r* | .60** | 0.34 | .45* |
| | *n* | 26 | 25 | 31 |

* $p < .05$. ** $p < .01$.

## Placement cut score

Binary logistic regression analysis is the preferred approach for estimating the probability of the presence or absence of an event (i.e., passing Chemistry 142) based on the value of a predictor variable or set of variables (i.e., placement test scores). We used this approach to determine the optimal test cut score for placement into Chemistry 142, conducting the analysis separately for each of three course outcomes (Chemistry 142 target course grade = 1.7, 2.0, and 2.3). The predictor variable was Chemistry 142 placement test score, and we set a criterion level of $p = .80$ as the probability of obtaining the target course grade. As shown in Table 8, the suggested cut scores obtained ranged from 12-18. If the goal is to insure that test-takers have a high probability of obtaining a Chemistry 142 grade of 1.7 or better, then the placement exam cut-score should be *at least 12* (out of 26). This was the minimum score across all three test versions and all three course sections for which the estimated probability met or exceeded .80.

**Table 8. Results from logistic regression analyses predicting CHEM 142 grade outcome from placement test score**

| | Target Grade = 1.7 | | | Target Grade = 2.0 | | | Target Grade = 2.3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Vers. A | Vers. B | Vers. C | Vers. A | Vers. B | Vers. C | Vers. A | Vers. B | Vers. C |
| **Predictor** | ***B*** **(SE *B*)** | ***B*** **(SE *B*)** | ***B*** **(SE *B*)** | ***B*** **(SE *B*)** | ***B*** **(SE *B*)** | ***B*** **(SE *B*)** | ***B (SE B)*** | ***B (SE B)*** | ***B (SE B)*** |
| **Test Score** | .13 (.06) | .24 (.07) | .26 (.07) | .19 (.05) | .28 (.06) | .23 (.06) | .21 (.04) | .33 (.05) | .22 (.05) |
| **Constant** | .34 (.98) | -1.25 (.99) | -1.27 (1.03) | -1.13 (.82) | -2.34 (.87) | -1.17 (.92) | -2.17 (.68) | -4.07 (.79) | -1.92 (.70) |
| **% earned grade** | 93 | 92 | 93 | 88 | 88 | 91 | 78 | 78 | 81 |
| **Test Score** | **Estimated probability of earning target grade** | | | | | | | | |
| **1** | .62 | .27 | .27 | .28 | .11 | .28 | .12 | .02 | .15 |
| **2** | .65 | .31 | .32 | .32 | .14 | .33 | .15 | .03 | .18 |
| **3** | .68 | .37 | .38 | .36 | .18 | .38 | .17 | .04 | .22 |
| **4** | .71 | .42 | .44 | .41 | .22 | .44 | .21 | .06 | .26 |
| **5** | .73 | .48 | .50 | .46 | .28 | .49 | .24 | .08 | .30 |
| **6** | .76 | .54 | .57 | .51 | .33 | .55 | .28 | .11 | .35 |
| **7** | .78 | .60 | .63 | .55 | .40 | .60 | .32 | .15 | .40 |
| **8** | .80 | .65 | .68 | .60 | .47 | .66 | .37 | .20 | .46 |
| **9** | .82 | .70 | .74 | .65 | .53 | .71 | .42 | .25 | .51 |
| **10** | .84 | .75 | .78 | .69 | .60 | .75 | .47 | .32 | .56 |
| **11** | .86 | .79 | .82 | .73 | .67 | .79 | .52 | .40 | .62 |
| **12** | .87 | .83 | .86 | .76 | .72 | .83 | .57 | .48 | .67 |
| **13** | .89 | .86 | .89 | .80 | .78 | .86 | .62 | .56 | .71 |
| **14** | .90 | .89 | .91 | .83 | .82 | .88 | .67 | .64 | .76 |
| **15** | .91 | .91 | .93 | .85 | .86 | .90 | .71 | .71 | .79 |
| **16** | .92 | .93 | .94 | .87 | .89 | .92 | .75 | .78 | .83 |
| **17** | .93 | .94 | .96 | .89 | .91 | .94 | .79 | .83 | .86 |
| **18** | .94 | .95 | .97 | .91 | .93 | .95 | .82 | .87 | .88 |
| **19** | .95 | .96 | .97 | .93 | .95 | .96 | .85 | .90 | .90 |
| **20** | .95 | .97 | .98 | .94 | .96 | .97 | .87 | .93 | .92 |
| **21** | .96 | .98 | .98 | .95 | .97 | .97 | .89 | .95 | .93 |
| **22** | .96 | .98 | .99 | .96 | .98 | .98 | .91 | .96 | .95 |
| **23** | .97 | .98 | .99 | .96 | .98 | .98 | .93 | .97 | .96 |
| **24** | .97 | .99 | .99 | .97 | .99 | .99 | .94 | .98 | .96 |
| **25** | .98 | .99 | .99 | .98 | .99 | .99 | .95 | .99 | .97 |
| **26** | .98 | .99 | 1.00 | .98 | .99 | .99 | .96 | .99 | .98 |

## SUMMARY

The UW Chemistry department, in collaboration with the Office of Educational Assessment, developed two parallel forms of a 26-item exam to place students into Chemistry 142. Development took place over the course of four academic quarters (Spring 2015 through Winter 2016) and was conducted in two phases: item writing and forms development, and forms validation.

Item writing and forms development included creation of a pool of 120-items that were formed into six 20-item tests and administered to enrolled Chemistry 142 students in Autumn 2015. Based on item analysis results, items were selected for inclusion in three parallel 26-item test versions that were piloted in Winter 2016.

Analyses of data from the Winter pilot included examination of test difficulty and internal consistency, item quality, the relationship of test scores to achievement indicators, and determination of a placement cut score. Based on results of these analyses, two parallel 26-item placement exams were created and first introduced into use in June 2016.