

Assessment of preparedness of first-year chemistry students: development and application of an instrument for diagnostic and placement purposes

Marietjie Potgieter¹, Bette Davidowitz² and Elsie Venter³

¹Department of Chemistry, University of Pretoria, South Africa

²Department of Chemistry, University of Cape Town, South Africa

³Centre for Evaluation and Assessment, University of Pretoria, South Africa

E-mail: marietjie.potgieter@up.ac.za; bette.davidowitz@uct.ac.za, elsie.venter@up.ac.za.

Abstract

Many universities in South Africa use alternative admissions tests together with results of the Grade 12 examinations for access or placement. These tests focus on academic literacy and mathematical skills and do not provide information about proficiencies in disciplines other than mathematics. The implementation of a new curriculum for Grades 10-12 in South Africa generates a need to monitor preparedness of first-year students to align first-year curricula as closely as possible to the content of the new syllabi. The need for a tailor-made instrument to assess preparedness for chemistry at the tertiary level was prompted by the lack of suitable tests with the appropriate focus, depth and coverage for application in the South African context. This paper describes the development and evaluation of a test instrument designed to assess and monitor baseline conceptual understanding in chemistry at the secondary-tertiary interface. The study was carried out with mainstream students at the universities of Pretoria and Cape Town ($N = 513$ and 258 , respectively). Data analysed using the Rasch measurement model confirmed that the majority of students from both cohorts was reasonably well prepared for chemistry at the tertiary level. The use of the instrument for diagnostic purposes was demonstrated. It was possible to identify gaps in the assumed pre-knowledge of students who have qualified for admission to chemistry at a tertiary level at their respective institutions. Inadequate pre-knowledge was most noticeable in the areas of chemical reactions and electrochemistry. In addition, the instrument also shows promise as a placement tool within programmes at a particular institution, i.e. in either mainstream or academic development programmes. Its predictive ability compares well with non-South African placement tests and with Grade 12 mathematics performance, the latter is widely used for this purpose in South Africa.

Introduction

The South African education system has undergone major changes during the past century which generated the need for monitoring of preparedness of first-year students and re-alignment of tertiary education to accommodate these changes. The Department of Education chose to adopt outcomes-based education as the foundation for the curriculum in South Africa (DoE, 2005). Because a relatively small percentage of learners enter tertiary education after graduating from secondary schools, the Department of Education has adjusted the focus of secondary education to satisfy a much wider range of constituencies than was the case in the previous dispensation.

The implementation of outcomes-based education as well as a new syllabus for physical sciences in the Further Education and Training phase of secondary schools in South Africa (Grades 10-12) (DoE, 1998) is likely to significantly alter the proficiencies of future first-year students. Tertiary

institutions would do well to anticipate and accommodate these changes in order to ensure a smooth transition from secondary to tertiary education. It will be the responsibility of tertiary educators to align first-year curricula as closely as possible to the new syllabi for their respective disciplines at secondary level. With the possible exception of academic development programmes, however, most university lecturers in physical sciences assume that students will have a reasonable basic knowledge and understanding of subjects such as mathematics, chemistry and physics. Such assumptions are seldom formally tested or confirmed. From a tertiary perspective, therefore, it is imperative that shifts in proficiencies of students upon entry to tertiary education are carefully monitored, especially during a transition period.

Preparedness for tertiary chemistry in our view does not consist only of knowledge, but also of conceptual understanding of fundamental concepts assumed as pre-knowledge for tertiary chemistry. In addition, an appropriate level of mathematical and process skills (understanding of chemical terminology and representational competence) is required. Representational competence refers to the ability to translate between different thinking levels and modes of representation (Ainsworth, 2006; Johnstone, 1991), a skill which is central to problem solving in chemistry. The purpose of this study is, therefore, the development of a test instrument for thorough assessment of preparedness at the secondary-tertiary interface to inform teaching and placement at the tertiary level. A literature search revealed the lack of suitable test instruments with the appropriate focus, depth and coverage for application in the South African context, which necessitated the development of a tailor-made instrument for our study. Results from the matriculation examination are not suitable for the diagnostic purposes that we propose. Firstly performance in chemistry is masked by the single comprehensive mark that is reported for both the physics and chemistry components of the syllabus; secondly, performance data on individual sections or questions in the physical sciences paper is not available. A further incentive for the development of our own instrument arises from concern over the use of matriculation results for selection and placement at tertiary institutions. Several authors have questioned the validity and fairness of the use of school-leaving results for such purposes, especially for students from ex-DET (Department of Education and Training) black-only education systems (Herman, 1995; Yeld & Haeck, 1997).

This paper describes the design and refinement of the test instrument and reports on its performance characteristics after application at the University of Pretoria (UP) and the University of Cape Town (UCT). As part of a larger ongoing project this instrument is used for biannual data collection at three universities that are representative of the tertiary landscape in South Africa (the Universities of Pretoria, Cape Town and Limpopo). The data will be used for the long-term monitoring of shifts in the proficiencies and skills of students entering tertiary chemistry for the first time.

Literature Review

Success at first-year level is a concern for tertiary institutions since funding depends on throughput of students in attaining their degree. Many studies have focused on selection and attrition at South African higher institutions (Eiselen & Geysers, 2003; Kahn, 1990; Lourens & Smit, 2003; Van der Vlier, Thijs & Zaaiman, 2003; Yeld & Haeck, 1997; Zaaiman, 1998, and Zaaiman, Van der Flier & Thijs, 2000). Various strategies are used for selection in higher education. Zaaiman (1998) reviewed several predictors of academic achievement, amongst them academic skills, learning potential, personality factors, gender and age. The first two of these predictors could be classified as academic factors. Academic skills are generally assessed using one of three types of tests, or a combination of these, namely achievement tests, ability tests and aptitude tests. Altink (1987) distinguishes carefully between these three constructs. She considers achievement and ability tests

to be at two different ends of the spectrum in terms of being subject or content bound. Achievement tests generally measure previously learnt subject knowledge while ability tests measure reasoning and other skills in a reasonably content-free environment. She places aptitude tests in the middle of this continuum as they make some use of previous knowledge, but are designed to assess subject-related skills, and comprehension and insight by using non-routine problems related to the course for which selection is being carried out. It is generally accepted that higher predictive validity in terms of success will be achieved from a combination of achievement and aptitude predictors than from only one of these (Zaaiman, 1998, and references cited therein).

Many universities in South Africa use alternative admissions tests, e.g. the Alternative Admissions Research Project (AARP) tests (AARP, 2007a & b) together with results of the Grade 12 examinations for access or placement or to assess students' learning needs. The Higher Education South Africa (HESA) National Benchmark Tests Project (Griesel, 2006) is an attempt to provide schools and higher education institutions with information about the competencies of students on entry to Higher Education by using benchmark tests in the areas of verbal reasoning, quantitative literacy and mathematical proficiency. These tests will not reproduce the essential information derived from the school-leaving examinations. The HESA initiative is expected to address broad issues of admission regulation and engage in the development of benchmark mechanisms within the higher education sector. It will, however, not yield subject specific assessment tools to inform chemistry teaching and placement at South African higher education institutions. We have chosen to design an aptitude test (comprehension, insight and skills) combined with some elements of achievement testing (factual recall of previous knowledge) for the joint purposes of diagnosis and prediction, i.e. to serve as an indicator of preparedness for first-year chemistry and to optimise its predictive ability for success in first-year chemistry.

The work by Van der Flier et al. (2003) suggests that the closer the focus of the testing is to the required performance, the better the correlation with performance will be. A literature survey of aptitude tests with good predictive power for academic success in first-year chemistry identified several sources of useful test items, but no single test instrument with a sufficient content coverage to also serve as a diagnostic tool was found. For example, the Chemical Concepts Inventory developed by Mulford and Robinson (2002) was designed to sample the extent of alternative conceptions in basic chemistry concepts, such as atoms and molecules, microscopic behaviour, heat and temperature, chemical formulas, gases, liquids and solutions, stoichiometry and phase changes. It consists of twenty-two conceptual questions in a multiple-choice format where known misconceptions were chosen as distractors. McFate and Olmsted (1999) designed a 25-item placement test for California State University Fullerton (CSUF) to replace the Toledo Chemistry Placement Exam that is commonly used for this purpose in the United States (Examinations Institute of ACS). The CSUF test was specifically designed to test skills levels rather than factual recall, but it was no more successful in minimising "incorrect" placements of students than comparable placement tests containing factual recall items. Wagner, Sasser and DiBiase (2002) developed a student pre-semester assessment instrument consisting of mathematics (10 items), chemistry (10 items) and demographic questions (8 items) to identify students at risk of failure in first-semester general chemistry. They showed that this instrument was the better predictor when compared to similar instruments, such as the Toledo examination, but that the predictive power of the instrument was not significantly improved by the inclusion of demographic variables. After pilot testing we selected the items from these instruments with the best characteristics as dictated by our objectives as will be described in a subsequent paragraph on methodology.

Research design

Shifts in proficiencies during a transition period and thereafter are expected to occur gradually. The appropriate research design is therefore a longitudinal study which makes provision for the standardisation of results to allow for direct comparison over an extended time period. Ideally, data should be collected at a range of tertiary institutions to allow generalisation of results to the majority of South African universities. While the South African education system is emerging from a period of in-depth renewal at primary and secondary level it is unlikely that similar far-reaching changes will be introduced at tertiary level. First-year syllabi in most SA tertiary institutions are fairly homogeneous and are not expected to undergo major changes in the near future. It will therefore be possible to implement a well-designed test instrument over an extended time period without modifications and still produce meaningful results for the tertiary community.

The objectives stated above for a longitudinal study can be met to a large extent by using the Rasch model (Bond & Fox, 2007) for statistical analysis of data. The Rasch model elevates test design to a level of sophistication and rigour that is not possible when using raw scores only, since this method generates *linear* item measures to reflect the relative difficulty of test items and *linear* person measures relating to the ability a person exhibits on a particular construct, in this case proficiency in chemistry. Applying the Rasch model enhances the quality of a project at various levels. The rigour of analysis is increased and the assumption of linearity is met – an assumption often ignored in the analysis of educational data (Boone & Rogan, 2005). In addition, person abilities and item difficulties can be aligned on the same linear scale using a unit called a logit (log-odds unit). This alignment allows evaluation of the quality of a test instrument in terms of the match between the range of item difficulties and the range of performance abilities of the respondents. The Rasch model also allows for equating of the test data between different versions of the same test as it evolves over a period of time (Bond & Fox, 2007). Comparison of results collected over many years, even if the instrument is subsequently modified, is therefore possible. The Rasch model is, therefore, ideally suited to the objectives of this study. The researchers plan to collect data every second year for a period of at least ten years in order to provide information about general trends and shifts in proficiencies.

Two tertiary institutions were selected for data collection in this study, namely the University of Pretoria (UP), a historically Afrikaans institution in Gauteng, and the University of Cape Town (UCT), a historically English institution in the Western Cape. These institutions differ in their location and in the traditional constituencies that they serve, but are both rated as first class tertiary institutions by international standards (Academic Ranking of World Universities, 2007). We intend to use results obtained at these institutions for benchmarking, i.e. to determine the minimum level of content knowledge and conceptual understanding of chemistry required for success at these institutions, which can be accepted as a general criterion for measuring preparedness for tertiary chemistry.

The factors taken into account for formulating design criteria for the test instrument will be considered next. The nature of teaching and assessment of the physical sciences at secondary level tends to be predominantly procedural in nature, while conceptual understanding is both required and nurtured in tertiary science education (Zoller & Tsaparlis, 1997). Anderson & Schönborn (2008) note that the term conceptual understanding used loosely in education practice, can be considered to be multi-faceted and “requires competence in the cognitive skills of memorization, integration, transfer, analogical reasoning, and system thinking” (Anderson & Schönborn, 2008, p309). Mayer (2002) describes meaningful learning and understanding as the ability to retain knowledge as well as the development of cognitive processes required to transfer that knowledge

into successful problem solving. He suggests that assessment tasks should require students to use their full range of cognitive processes; thus it is important to broaden the manner in which learning is assessed to probe whether students are capable of actively using their stored content knowledge in solving problems. From a tertiary perspective it is, therefore, imperative to design an assessment tool to probe content knowledge as well as conceptual understanding of chemistry.

The test instrument was intended to serve the dual purpose of diagnosis and prediction. For diagnostic purposes coverage of fundamental concepts generally accepted as pre-knowledge in all subject topics included in a typical first-year chemistry syllabus was required (content validity). In order to obtain a clear picture of the contribution of the new Physical Sciences syllabus to shifts in proficiencies, students who did not receive their secondary education in South Africa according to the South African education system were excluded from this study. Items assessing basic mathematical skills were included, since mathematical aptitude is well documented to be one of the strongest predictors of success in first-year general chemistry (McFate & Olmsted, 1999; Wagner et al., 2002). The best items for this purpose would be those that could be answered without the use of a calculator, rather than items requiring complex computations (see Wagner et al., 2002 for a similar strategy). Further requirements for the instrument included assessment of representational competence *i.e.* the ability to interpret symbolic, macro and submicro representations (Ainsworth, 2006; Johnstone, 1991) and assessment of conceptual understanding rather than recall or application of practiced procedures (Bowen & Bunce, 1997; Mayer, 2002). To this end, it would be important to choose *conceptual items* that tap a student's understanding of chemical ideas associated with a question, rather than *algorithmic questions* that can be answered by applying a set procedure to generate a response.

Methodology

Development and refinement of the test instrument

The majority of test items included in this instrument were obtained from sources in the literature. Several items were included from a multiple-choice test instrument developed by Mulford and Robinson (2002) for the determination of the extent of students' alternative conceptions about specific core first-year chemistry topics. Conceptual questions on the mole concept (Krishnan & Howe, 1994), chemical equilibrium (Huddle, 1998), solution chemistry (Smith & Metz, 1996), electrochemistry (Ogude & Bradley, 1994; Oswego Regents Exam Prep) and organic chemistry (Oswego Regents Exam Prep) were included. The items with best correlation with student performance from the Student Pre-Semester Assessment test (Wagner et al., 2002) and the CSUF Placement test (McFate & Olmsted, 1999) were also included. The first version of the test instrument was piloted in 2003 with several groups of Grade 12 learners at high schools in Pretoria. Content validity was established by a group of experts in secondary and tertiary education (Potgieter, Rogan & Howie, 2005).

After processing the results of the pilot test a first cycle of refinement was performed. Items with poor characteristics were either replaced or modified to generate the second version of the test instrument. This version of the instrument was used to collect data for several cohorts of first-time entering chemistry students at UCT and UP at the start of the 2004 academic year (Potgieter, Davidowitz & Blom, 2005). After interpretation of this data minor adjustments were made to existing test items to avoid ambiguity and to strengthen weak distractors, and a number of new items were added to ensure adequate coverage of assumed pre-knowledge for first-year chemistry. This process generated a third version of the test instrument for implementation early in 2005. It consisted of 75 test items that were divided into two similar subtests (Test A and Test B) in order to restrict the time required to administer the test. Test A consisted of 41 items and Test

B of 42 items. Both tests included 8 anchor items that were common to both subtests to allow for the equating of the subtests and subsequent merging of data.

In 2005 data was collected at the start of the academic year at UP and UCT. A final round of refinement was performed using both classical test theory and the Rasch model to evaluate the performance of each individual test item. It resulted in the exclusion of 10 items from the instrument. The final selection of 65 items showed excellent item characteristics, i.e. good item-test total correlation, good spread across different subtopics and appropriate range in the level of difficulty. The majority of these test items was in a multiple-choice format (60 items, 92%). The remainder of the items included short answer questions and a self-constructed submicroscopic representation which were assessed and coded by the examiners (5 items, 8%). Sixteen of the 65 test items were paired to provide an opportunity for explanation of answers in a follow-up question (two-tier methodology, Treagust, 1988). A significant number of schematic or pictorial representations were used (14 items; 12 macroscopic and submicroscopic representations plus two sets of structures of organic molecules). Examples of test items that closely resemble those included in the instrument are shown in Appendix A of this paper. Item E05 is a multiple-choice item with submicroscopic representations which require conceptual understanding as well as representational competence for their interpretation. It is paired with item E06 which requires a short explanation for the answer provided in item E05. For the majority of paired or two-tier items the second item provided possible explanations in multiple-choice format (not shown in Appendix A). In item B09, which assesses process skills, the problem is stated in scientific terminology and requires translation to a mathematical formula. Item K05 probes for mathematical skills related to concentration. All of the items in the Appendix A, with the possible exception of item K05, require conceptual understanding rather than rote learning and can therefore be considered to be conceptual questions.

The 65 test items of the final version of the test instrument were grouped into 11 subsets covering basic chemistry topics, special content topics and two categories of scientific process skills (see Table 1). A small number of items qualified for inclusion into more than one of the subsets, e.g. item K05 was included in three subtopics, the mole concept, solutions and mathematical skills. The test items were divided into two similar subtests (Test A and Test B), each consisting of 37 items including 9 common anchor items, with the objective to produce one set of measures by statistical methodology (Linacre, 1997). The test was considered to be stable and will be used without any further modification for data collection over a period of four to eight years. This paper reports only on data collected at the start of 2005 using the 65 items selected for the final version of the instrument.

Item analysis based on raw scores was performed throughout the development phase to inform the selection and refinement of items. The data were cleaned and prepared for analysis using SAS[®]. SPSS[®] was used to generate statistics typically associated with the classical test theory such as frequencies and reliability coefficients. The corrected item-test total correlation statistic was used as a discrimination value for item selection. In the final phase of refinement of the test instrument the authors took advantage of the strength of the Rasch model to verify and substantiate decisions taken earlier during test development.

Sample

The sample used for data collection at the start of the academic year in 2005 consisted of two cohorts of first-year mainstream students at the University of Pretoria ($N = 513$) and the University of Cape Town ($N = 258$). Mainstream students are those first-year students who meet entrance requirements to the Science Faculty at their particular university. The sample excluded

students who completed high school outside South Africa or sat for international school leaving examinations. In addition, students repeating the course did not participate in the study. The difference in sample size at the two institutions can be explained as follows: not only does UCT admit fewer students due to a smaller capacity, but their first-year cohort consists of a larger component of students with school-leaving qualifications from institutions other than the South African Department of Education and who were therefore excluded from this study.

Findings

Comparison of raw score results

Based on the answers obtained for short answer test items during the piloting and refining of the test instrument the researchers have formulated a set of typical responses for coding purposes. Only one of these responses for each test item was accepted as the correct answer. It became clear that three of the short answer test items (two requiring an explanation similar to item E06 in Appendix A, and the third requiring a submicroscopic representation) were especially demanding in terms of the insights required for a correct answer to be produced. It was, therefore, possible for a respondent to display legitimate conceptual understanding without being able to generate an answer that is scientifically accurate. In such cases partial credit was awarded. Correct, partially correct and incorrect answers were assigned scores of 1.0, 0.5 and 0.0, respectively.

The performance of different cohorts of students as reflected by raw score data is presented in Table 1. The *performance index* calculated per subtopic for each cohort of students is the average of the percentages of correct answers per test item in the subset (raw scores). The standard errors in Table 1 reflect the range of difficulties for the items in a particular subset. For example, the mean of 65.8% for the subset Atoms and Ions for the UP cohort was calculated from an average of correct answers ranging between 22.4 and 85.6% for the questions in that subset.

Table 1 Performance indices for the mainstream cohorts (2005 data collection)

Topic (Subset)	N (items)	UP (N = 513)	Std error	UCT (N = 258)	Std error
<i>Basic Concepts</i>					
Atoms & Ions	8	65.8	8.7	68.1	7.4
Mole concept	6	52.4	10.1	58.8	11.2
Phases of matter	8	56.8	6.2	65.2	5.9
Solutions	6	51.7	9.1	56.6	10.0
Reactions	8	35.8	7.6	41.4	6.1
<i>Special Topics</i>					
Acids & Bases	6	65.6	6.2	58.7	6.8
Chemical equilibrium	6	53.5	8.0	51.4	8.9
Electrochemistry	5	39.6	6.4	44.0	8.7
Organic chemistry	6	58.0	8.1	53.3	8.7
<i>Process skills</i>					
Skills (Maths)	10	51.8	8.2	60.0	7.3
Skills (Comprehension)	6	46.1	8.5	53.3	8.8

In general, the UCT mainstream cohort performed better than the UP mainstream cohort on all of the subtopics dealing with basic concepts and process skills. For example, the UCT performance

index for Phases of Matter (65.2%, standard error = 5.9) is higher than that of the UP cohort (56.8%, standard error = 6.2). Similarly, the UCT performance index for Skills (mathematics) (60.0%, standard error = 7.3) is higher than that of the UP cohort (51.8%, standard error = 8.2). The trend is reversed for some of the special topics, such as Acids & Bases (UCT: 58.7% vs. UP: 65.6%). A comparison of Grade 12 final examination results for the two cohorts is shown in Table 2. The performance in Grade 12 physical science of students in the two cohorts is almost identical as judged by the frequency distributions reported in Table 2, but the UCT cohort performed better in Grade 12 mathematics. The performance difference between the two cohorts as expressed in terms of Grade 12 mathematics corresponds more closely to their performance in our test than their Grade 12 physical science results, which is in line with the findings of McFate and Olmsted (2002) that mathematics background is a better predictor for success in first-year chemistry than chemistry knowledge.

Table 2 Performance of mainstream cohorts in Grade 12 final examination^a

	Composition of cohorts according to Grade 12 matric results ^b					
	% HG A	% HG B	% HG C	% HG D	% HG E	SG A & B (%)
UCT: Mathematics	39	21	22	10	3	5
UP: Mathematics	25	20	29	23	2	1
UCT: Physical Science	26	27	30	13	3	1
UP: Physical Science	26	25	35	13	<1	<1

^a Excluding students who sat for international school leaving examinations

^b HG = higher grade, SG = standard grade

A thorough analysis of response frequencies for individual items in each subtopic highlighted a number of aspects with direct implications for teaching at first-year level. These results have been reported elsewhere (Potgieter, Davidowitz & Blom, 2005), but will be summarised here. Students lack a sound foundation in chemical reactions (stoichiometry, conservation of mass, dissociation of salts in aqueous medium, and the energetics of chemical bonding) and their preparation for electrochemistry is so weak that lecturers cannot assume any pre-knowledge of this topic. Students struggle to carry out manipulations of algebraic expressions containing exponents, a factor that must be taken into account, especially when teaching topics such as stoichiometry, gas laws, acids and bases, and chemical equilibrium. Results also show serious inadequacies in the ability of students to translate between different thinking levels and modes of representation (Ainsworth, 2006; Ainsworth, 2006) which is central to an understanding of chemical concepts.

Rasch model results

Since students in the mainstream cohorts at UCT and UP had met the entrance criteria for the respective science faculties of these two first class South African institutions (Academic Ranking of World Universities, 2007), it was assumed that they had the necessary pre-knowledge for chemistry at the tertiary level. The data collected for the two groups from UCT and UP were combined ($N = 770$) in order to calculate item difficulties through application of the Rasch measurement model. These item difficulties can be utilised as a benchmark to compare shifting trends in students' performance in the future. This means that the item measures (difficulty values) of specific "anchor" items will be used to compare future results to fixed points of reference. Fluctuation in performance can then be determined because the item difficulty values have been predetermined. The software package, WINSTEPS, was used to generate the item difficulty and person ability measures. Person ability measures are plotted directly against item difficulty measures on the same scale, because they are estimated in the same metric. The item mean was set or centred at 0. The "person-item map" is shown in Figure 1.

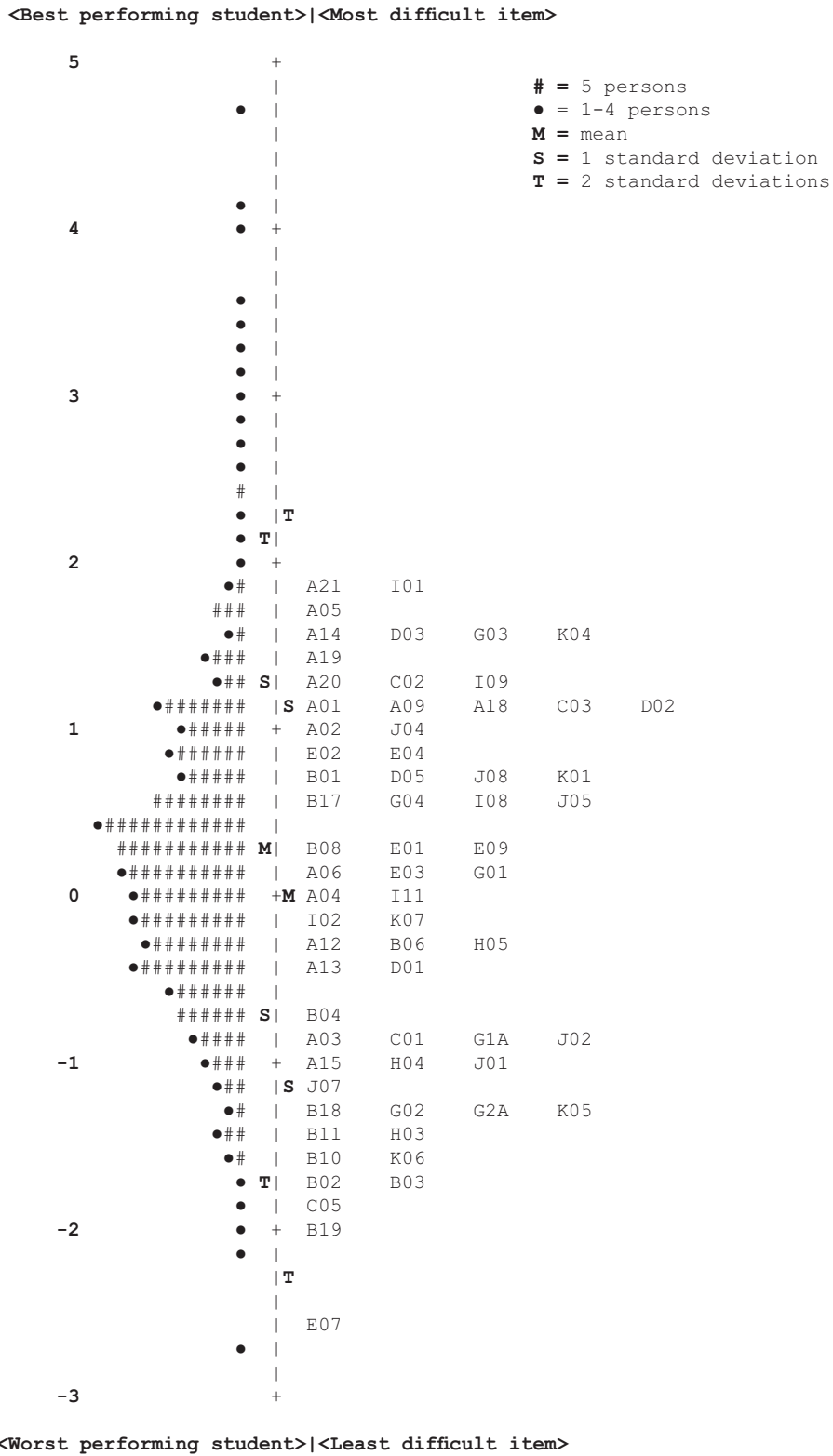


Figure 1 Person-item map of combined data for mainstream cohorts at UCT and UP. See Appendix B for details on item codes.)

The vertical line in Figure 1 separates the item difficulties (on the right) and the person abilities (on the left). The scale is from -3 to 5 logits, where -3 is the lowest and 5 the highest measure. This corresponds to the easiest and the most difficult items, respectively. A logit (log-odds unit) is a unit of interval measurement (Boone & Rogan, 2005). For a student (person) with an ability of 0 logits on the scale, the probability for that student to correctly answer items of a corresponding difficulty is 50%. In other words, considering Figure 1, such a student will have a 50% chance of correctly answering item A04 or item I11. The same student will only have a 25% probability of correctly answering item A02 or item J04 (difficulty of 1 logit). The probability of this student answering items A15, H04 or J01 (difficulty of -1 logit) correctly, is 75%. As shown in Figure 1 items A21 and I01 are the most difficult items and Item E07 is the easiest. There is a reasonable spread of items in terms of difficulty, with only two small gaps at logits +0.45 and -0.60. None of the items achieved a difficulty level of greater than 2 logits, but a number of person measures above 2 logits were recorded. All the items, except E07 lie within two standard deviations from the mean. The mean, M , of the person abilities on the left of the vertical line is approximately 0.2 logits higher than the mean for the item measures. This indicates that the test was on average slightly easy for the majority of the students. Students with an ability higher than 2 logits, approximately 4% of students, found the test too easy and had a 100% probability of answering every item correctly.

The range of difficulties of test items will be considered next. In general, there was a good match between the range of difficulties of test items (item measures) and the performance range (person measures) for the benchmark cohorts (Figure 1). Three scenarios were found for the different subtopics comprising the test instrument, namely subtopics with an even balance between easy and difficult items and subtopics with a bias towards either easy or difficult items. The subtopics, Mole Concept, Solutions, Organic Chemistry and Skills (Language and others) were evenly balanced (refer to Appendix B for codes and a description of items included in each subtopic). Subtopics Atoms & Ions, Phases of Matter and Acids & Bases were slightly biased towards easy items. The subset, Skills (mathematics) was marginally biased towards difficult items while subtopics Electrochemistry, Chemical Equilibrium and Reactions contained a majority of difficult items. The raw scores gave a similar picture (Table 1), but the Rasch method uncovers the subtleties described above. It should be emphasized that item difficulties are based on raw score performance and as such reflect the extent to which the combined cohorts of mainstream students at UP and UCT struggled to provide correct answers to test items. Subsets biased towards difficult items cover content areas for which students are not adequately prepared. As a result they reflect the gaps in the assumed pre-knowledge of students who have qualified for admission to chemistry at a tertiary level.

The distribution of person measures for the mainstream cohorts of the two institutions involved in this study is plotted as the number of students per logit interval, see Figure 2. This graphical representation provides a powerful tool to compare the ranges of person abilities for different cohorts of students. Figure 2 shows that the majority of students at both institutions are reasonably well prepared for tertiary chemistry (they have person measures larger than 0.0 logits). Similar representations of person measures may be used to inform or evaluate the placement of students in different cohorts at the same institution, i.e. either in the mainstream cohort or in an academic development programme that is offered for under-prepared students. For placement purposes it would be important to determine the threshold for success in first-year chemistry, i.e. a cut-off score for person measures below which there is a high probability of failing. We have arbitrarily taken a logit of 0.0 as a reference point for preparedness for chemistry, but the validity of this choice will have to be determined in future. In a separate study the use of this instrument for the evaluation of placement procedures at three South African institutions was investigated (Potgieter,

Davidowitz & Mathabatha, 2008). A substantial overlap between the ranges of proficiencies of students in the mainstream and academic development programme at the University of Limpopo was found. It was clear that the majority of students placed in the mainstream of that institution were no better prepared than those in the development programme.

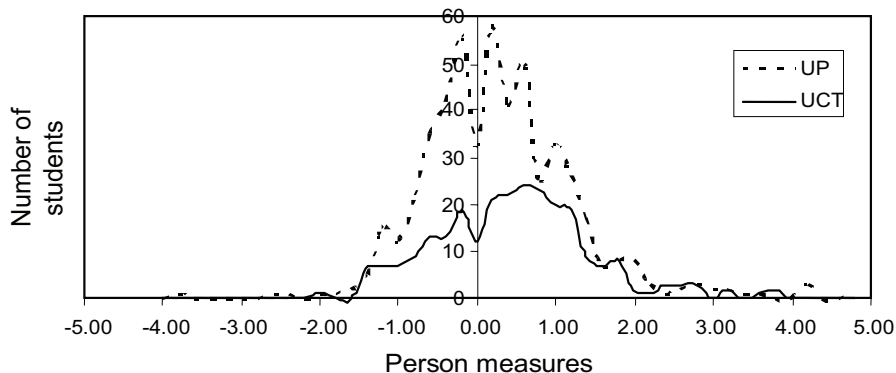


Figure 2 Person measures for the mainstream cohorts at the Universities of Pretoria (UP) and Cape Town (UCT).

The internal consistency of the subtests of the instrument was determined for each sample group. The Cronbach alpha coefficients (Cronbach, 1951) computed with SPSS software® gave values between 0.80 and 0.83 for the mainstream cohorts of UP and UCT, respectively. The Winsteps item reliability was 0.99 and the person reliability was 0.81. The high item reliability indicates an appropriate range of item difficulties while the high person reliability, in turn, is indicative of a good range of person abilities on the underlying construct. The construct 'reliability', as reflected by Cronbach alpha values in the classical test theory setting, pertains to the reproducibility of test results. This construct is equivalent to Rasch's 'person reliability'. An 'item reliability', which considers the 'true item variance/observed item variance', is also reported by Winsteps (Winsteps® & Facets Rasch Software). Generally, 'Rasch reliability is more conservative ...' than classical reliability (Institute for Objective Measurement). Fit statistics for the test have been examined, a small number of misfitting persons were removed and after final refinement of the instrument, all items fit the model. Based on these statistical parameters the test instrument was judged to have performed according to expectations and will therefore remain unchanged during future rounds of implementation.

Correlations between different performance indicators

In order to judge the potential of the test instrument for selection and placement a number of correlations were calculated. These include the correlations between person measures, Grade 12 mathematics (higher grade) results, and chemistry performance at first-year level, respectively (Table 3). The predictive power of performance in Grade 12 mathematics for performance in first-year general chemistry has been well documented (McFate & Olmsted, 1999; Wagner et al., 2002) and it is widely used by South African tertiary institutions to inform admission and placement. The chemistry performance results used for the calculation of correlations are different for the two institutions, because first-year chemistry instruction at the University of Pretoria is divided into two semester modules whereas a single course in chemistry is offered during the first-year at UCT. Group sizes reported in Table 3 differ marginally from those reported in Table 1 due to incomplete records of Grade 12 performance data and student attrition during the first year.

The results reported in Table 3 indicate that there is a moderate correlation of person measures with academic performance in mainstream first-year chemistry. This correlation is similar to the correlation between performance in Grade 12 mathematics and Semester 1 chemistry at UP, but significantly weaker than the correlation between performance in Grade 12 mathematics and Year 1 chemistry at UCT. It compares favourably, however, with correlations of 0.30-0.60 between course grades and pre-assessment scores reported in the literature (Wagner et al., 2002). The higher correlation between performance in Grade 12 mathematics and Year 1 chemistry at UCT can at least in part be ascribed to the fact that prediction of success is typically more accurate for better performing students. The mainstream UCT cohort comprises a higher percentage of better performing students, see Table 2.

Table 3 Correlations between performance indicators

	Grade 12 higher grade mathematics ^a	Person measures
UP mainstream Semester 1 results	0.487 (<i>N</i> = 467)	0.563 (<i>N</i> = 467)
UCT mainstream Year 1 results	0.716 (<i>N</i> = 212)	0.561 (<i>N</i> = 243)

^a Excluding students who sat for international school leaving examinations

The correlation of person measures with academic performance in first-year chemistry shown in Table 3 is statistically significant at the 0.01 level and an indication of the predictive power of the instrument which represents the middle-ground resulting from the dual-purpose design of the instrument, i.e. that it is intended for both diagnosis and placement. One possible cause of the lower correlation and predictive power may be that our instrument probes deeper understanding of chemistry while the assessment in first-year chemistry might focus more on memorisation and reproduction of facts and less on understanding. A selection of test items could conceivably be made from this instrument for significantly enhanced predictive power, but at the expense of content coverage. Further work is required to establish the specificity and sensitivity of the instrument (Wagner et al., 2002) and to determine whether similar results are obtained during later cycles of implementation.

Conclusions and Implications

The test instrument described in this paper can be used for diagnostic purposes because of its specific design to probe conceptual understanding over the complete range of topics that are typically included in first-year chemistry syllabi in South Africa. It has been shown to highlight several content areas, e.g. chemical reactions and electrochemistry, where the foundation from secondary education is inadequate. This creates the possibility for articulation with science education at secondary or tertiary level to correct misconceptions and to encourage the development of essential skills for the study and mastery of chemistry. In another paper (Potgieter, Davidowitz & Blom, 2005) we have shown serious inadequacies in the representational competence of students and have warned that students lack essential mathematical skills for tertiary chemistry.

The combination of this test instrument for data collection with the Rasch method for statistical analysis provides a powerful tool for placement purposes. The rigour of the Rasch method contributes to the validity of person measures as a reflection of preparedness for chemistry education at South African tertiary institutions. The correlation of 0.56 between person measures and first-semester or first-year performance in chemistry at UP and UCT, respectively, compares favourably with correlations of 0.30-0.60 between course grades and pre-assessment scores reported in the literature (Wagner et al., 2002). Person measures can be used to inform decisions

regarding placement within specific programmes at the same institution, i.e. in either mainstream or academic development programmes. In addition, the complete test instrument or specific items with high predictive power can be included in the battery of selection tests currently used by tertiary institutions. Further work is required to determine the ability of the instrument to correctly predict passing or failing in first-year chemistry and to determine the cut-off score of person measures below which a student should be considered at risk of failing.

Ground work has now been completed for the monitoring of shifts in conceptual understanding and skills development of university entrants as a result of the introduction of the new Physical Sciences syllabus in secondary education. The instrument described in this paper will be used by the authors over a period of ten years to collect data at three or more tertiary institutions representative of the tertiary landscape in South Africa to allow for the generalisation of results. The information that is gathered will enable tertiary educators to accommodate changes in proficiencies and to align first-year curricula as closely as possible so that a smooth transition from secondary to tertiary education can be achieved. It can also be used to inform curriculum and syllabus refinement at secondary level.

Acknowledgements

This work is based upon research supported by the National Research Foundation of South Africa (NRF). Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and therefore the NRF does not accept any liability in regards thereto.

References

- Academic Ranking of World Universities, Graduate School of Education, Shanghai Jiao Tong University, August 2007. Retrieved 22 August, 2008 from <http://www.arwu.org/rank/2007/ranking2007.htm>.
- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction* 16(3), 183-198.
- Alternative Admissions Research Project (AARP) (2007a). The Placement Test in English for Educational Purposes (PTEEP): The Tea Test. AARP, Academic Development Programme, University of Cape Town.
- Alternative Admissions Research Project (AARP) (2007b). The Mathematics Comprehension Test (MCOM). AARP, Academic Development Programme, University of Cape Town.
- Altink, W.M.M. (1987). The evaluation of selection tests for educational upgrading programmes in Botswana and Swaziland. *International Journal of Educational Development*, 7(1), 1-12.
- Anderson, T.R. and Schönborn, K.J. (2008). Bridging the Educational Research-Teaching Practice Gap: Conceptual understanding, Part 1: The multifaceted nature of expert knowledge, *Biochemistry and Molecular Biology Education*, 36(4), 309-315.
- Bond, T.G. & Fox, C.M. (2007). Applying the Rasch model. New Jersey: Lawrence Erlbaum Associates.
- Boone, W. & Rogan, J. (2005). Rigour in quantitative analysis: The promise of Rasch analysis techniques. *African Journal of Research in Mathematics, Science and Technology Education*, 9(2), 25-38.
- Bowen, C.W. & Bunce, D.M. (1997). Testing for conceptual understanding in general chemistry.

Assessment of preparedness of first-year chemistry students: development and application of an instrument for diagnostic and placement purposes

The Chemical Educator, 2(2), 1-17.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

Department of Education (DoE) (2005). *National Curriculum Statement Grades 10-12 (General) Learning Programme Guidelines*. Retrieved 22 August, 2008 from http://curriculum.pgwc.gov.za/php/circular_docs/55_maths.pdf

Department of Education (DoE) (1998). *Education White Paper 4: A Programme for the Transformation of Further Education and Training*. Government gazette, Vol. 399, No. 19281. Pretoria: Government Printer

Eiselen, R. & Geysers, H. (2003). Factors distinguishing between achievers and at risk students: a qualitative and quantitative synthesis. *South African Journal of Higher Education*, 17(2), 118-130.

Examinations Institute of the American Chemical Society Division of Chemical Education. Retrieved 22 August, 2008 from <http://www4.uwm.edu/chemexams/>.

Griesel, H. (Ed). (2006). Access and Entry Level Benchmarks. The National Benchmark Tests Project. Pretoria: Higher Education South Africa – HESA.

Herman, H. D. (1995). School-leaving examinations, selection and equity in higher education in South Africa, *Comparative Education*, 31(2), 261-272.

Huddle, B.P. (1998). "Conceptual Questions" on LeChâtelier's principle. *Journal of Chemical Education*, 75(9), 1175.

Institute for Objective Measurement, Inc. website. Retrieved 22 August, 2008 from <http://www.rasch.org/>.

Johnstone, A.H. (1991). Why is Science Difficult to Learn? Things Are Seldom What They Seem. *Journal of Computer Assisted Learning*, 7, 75-83.

Krishnan, S.R. & Howe, A.C. (1994). The mole concept. *Journal of Chemical Education*, 71(8), 653-655.

Kahn, M.J. (1990). Some questions concerning the standard and role of external examinations *Studies in Educational Evaluation*, 16(3), 513-527.

Linacre, J.M. (1997) KR-20 or Rasch Reliability: Which Tells the "Truth"? *Rasch Measurement Transactions*, 11:3 580-581. Retrieved 22 August, 2008 from <http://www.rasch.org/rmt/rmt1131.htm>.

Lourens, A. & Smit, I.P.J. (2003). Retention: predicting first-year success. *South African Journal of Higher Education*, 17(3), 169-176.

McFate, C., & Olmsted III, J. (1999). Assessing student preparation through placement tests. *Journal of Chemical Education*, 76(4), 562-565.

R. E. Mayer (2002) Rote versus meaningful learning, *Theory into Practice*, 41(4), 226–232.

Mulford, D.R. & Robinson, W.R. (2002). An inventory for alternate conceptions among first-semester General Chemistry students. *Journal of Chemical Education*, 79(6), 739-744.

Ogude, A.N. & Bradley, J.D. (1994). Ionic conduction and electrical neutrality in operating electrochemical cells. *Journal of Chemical Education*, 71(1), 29-34.

Oswego city school district Regent exam prep center, Chemistry section. Retrieved

22 August, 2008 from <http://regentsprep.org/Regents/core/questions/topics.cfm?Course=CHEM>.

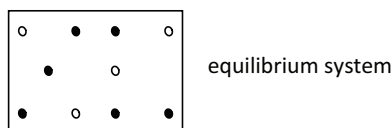
- Potgieter, M., Davidowitz, B. and Blom, B. (2005). *Chemical Concepts Inventory of First Year Students at Two Tertiary Institutions in South Africa*. Proceedings of the conference of the South African Association of Research in Mathematics, Science and Technology Education, Namibia, January 2005, pages 664-675.
- Potgieter, M., Davidowitz, B. & Mathabatha, S.S. (2008). Preparedness for tertiary chemistry: issues of placement and performance of academic development programmes. *South African Journal of Higher Education*, in press.
- Potgieter, M, Rogan, J.M. & Howie, S. (2005). Chemical concepts inventory of Grade 12 learners and UP foundation year students. *African Journal for Research in Mathematics, Science and Technology Education*, 9(2), 121-134.
- Smith, K.J., & Metz, P.A. (1996), Evaluating student understanding of solution chemistry through microscopic representations. *Journal of Chemical Education*, 73(3), 233-235.
- Treagust, D.F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159-169.
- Van der Flier, H., Thijs, G.D. & Zaaiman, H. (2003). Selecting students for a South African mathematics and science foundation programme: effectiveness and fairness of school-leaving examinations and aptitude tests. *International Journal of Educational Development*, 23(4), 399-409.
- Wagner, E.P., Sasser, H., & DiBiase, W.J. (2002). Predicting students at risk in General Chemistry using pre-semester assessment and demographic information. *Journal of Chemical Education*, 79(6), 749-755.
- WINSTEPS® & Facets Rasch Software. (2007). Retrieved 22 August, 2008 from <http://www.winsteps.com>.
- Yeld, N. & Haeck, W., 1997. Educational histories and academic potential: can tests deliver? *Assessment and Evaluation in Higher Education*, 22(1), 5-16.
- Zaaiman, H., 1998. Selecting Students for Mathematics and Science. The Challenge Facing Higher Education in South Africa. HSRC, Pretoria.
- Zaaiman, H., Van Der Flier, H., & Thijs, G.D. (2000). Selection as contract to teach at the student's level. Experiences from a South African mathematics and science foundation year. *Higher Education*, 40, 1-21.
- Zoller, U, & Tsaparlis, G. (1997). Higher and lower-order cognitive skills: The case of chemistry. *Research in Science Education*, 27(1), 117-130.

Appendix A: Examples of test items

Note: Some of the examples shown below do not appear in the final version of the test instrument.

Examples 1 and 2: Multiple-choice item with submicroscopic representations followed by a short answer item requiring a self-constructed explanation.

- E05. The exothermic reaction $\circ(g) \rightleftharpoons \bullet(g)$ was allowed to come to equilibrium, as represented in box below:



The pressure of the system at equilibrium was increased without affecting the temperature. Which box (A – E) best represents the composition of the mixture when the new equilibrium position is reached?

Box A	Box B	Box C	Box D	Box E

- E06. What is the reason for your answer to the previous question?

Example 3: Probing for process skills (ability to interpret scientific terminology, translation between different ways of symbolic presentation)

- B09. Coulomb's law states that force of attraction (F) between two charged particles is inversely proportional to the square of the distance between them (d). The symbol k is a constant. What is the formula of the equation described by this statement?

(a) $F = kd^2$ (b) $\frac{k}{F^2} = d$ (c) $F = \frac{k}{\sqrt{d}}$ (d) $F = \frac{k}{d^2}$ (e) $\sqrt{F} = dk$

Example 4: Probing for mathematical skills

- K05. What is the total number of moles of H_2SO_4 needed to prepare 5,5 litres (dm^3) of a 2,0 $mol \cdot dm^{-3}$ solution of H_2SO_4 ?

- (a) 2,25 (b) 5,5 (c) 11 (d) 22

Appendix B: Description of test item codes included in each subtopic

Codes in brackets refer to the format of items: MC = multiple-choice, SA = short answer, calc = calculation required for answer, macro = macroscopic representation(s), and micro = submicroscopic representation(s) included in the problem statement. Two-tier items are shown as A12 + A13, for example.

Atoms & Ions:

A14(MC, micro), B06(MC), B10(MC), B17(MC), B19(MC), C01(MC), C05(MC), K06(MC)

Mole concept:

B02(MC, calc), D01(MC), D02(MC), D03(MC), D05(SA, calc), K05(MC, calc)

Phases of matter:

A02(MC), A03(MC), A06(MC, micro), A12(MC, macro) + A13(MC), H03(MC), H04(MC), K01(MC, micro)

Solutions:

A04(MC), A15(MC, micro), A20(MC, macro) + A21(MC), H05(MC), K05(MC, calc)

Reactions:

A01(MC), A05(MC, micro), A09(MC), A18(MC) + A19(MC), B18(MC), G03(SA, micro required), K07(MC)

Chemical equilibrium:

E01(MC, micro) + E02(SA), E03(MC, micro) + E04(SA), E07(MC), E09(MC)

Acids & Bases:

G01(MC, micro) + G1A(MC), G02(MC, micro) + G2A(MC), G04(MC), K07(MC)

Electrochemistry:

I01(MC), I02(MC), I08(MC), I09(MC), I11(MC)

Organic chemistry:

J01(MC), J02(MC), J04(MC), J05(MC), J07(MC), J08(MC)

Skills (Mathematics):

B01(MC, calc), B02(MC, calc), B03(MC, calc), B08(MC, calc), B11(MC, calc), C02(MC, calc), C03(MC, calc), D05(SA, calc), K05(MC, calc), A14(MC, micro)

Process Skills (Language & others):

B04(MC), C01(MC), K04(SA), H05(MC), G03(SA, micro required), B17(MC)