The Design of an Algebra Concept Inventory

Diarmaid Hyland[1], Ann O'Shea[1], Sinéad Breen[2], Brien Nolan[2] and Paul van Kampen[2]

[1]Maynooth University and [2]CASTeL, Dublin City University.

In this paper, we report on the development of an Algebra Concept Inventory (ACI) intended for students transitioning from second- to third-level education. We begin by outlining the work done on concept inventories to date before describing some of the guiding principles of interdisciplinary design teams. Our methodology for developing the ACI is detailed in a step-by-step manner including the formation of the design team, defining the parameters of the ACI, shortlisting items, and piloting. The iterative design process resulted in a 31-item preliminary ACI. We conclude by highlighting aspects of how the interdisciplinary team functioned throughout the process before outlining some potential uses of the ACI which was rolled out during Semester 2 of the 2020-2021 academic year to all higher education institutions (HEIs) in Ireland. The ACI is available for interested practitioners upon request.

Introduction and Background

In this paper, we outline the creation of an Algebra Concept Inventory (ACI) with a focus on how the interdisciplinary design team was formed and collaborated throughout the process.

What is a Concept Inventory?

A concept inventory (CI) is a set of questions designed to assess students' conceptual understanding of a given topic. They originated in Physics Education Research in the early 1990s with the Force Concept Inventory (FCI) (Hestenes et al., 1992) which was based on the Mechanics Diagnostic Test (Halloun & Hestenes, 1985). The success of the FCI in physics has seen CIs designed in other subjects. Within mathematics, CIs have been designed and implemented in the areas of function (O'Shea et al., 2016), calculus (Epstein, 2013), precalculus (Carlson et al., 2010), and statistics (Stone et al., 2003). We believe that the work done on building CIs in mathematics is of huge value to the community. CIs have multiple applications in research and instruction. Though our immediate motivation for developing an ACI is as a diagnostic tool, CIs can be used for evaluating instruction, and as a placement examination (Hestenes et al., 1992). Thus, we seek to build on what has already been done by developing a concept inventory for algebra, which is fundamental to virtually all tertiary level mathematics and is a prerequisite to the aforementioned topics.

Interdisciplinary Teamwork

A key aspect of this work is creating an interdisciplinary team that will work together to develop the ACI. It is common for interdisciplinary teams to be formed to undertake projects such as this, where a certain topic is relevant to several adjacent areas of expertise (Claus & Wiese, 2019). In our case, we are developing a research tool with relevance to members of the mathematics and education communities, as well as educators in other adjacent subjects. Barriers to interdisciplinary team-based research are well described (e.g., Morse et al., 2007) and are often logistical; COVID-19 has exacerbated these problems. For this project to progress as effectively as possible, research on interdisciplinary teamwork was examined for guiding principles and best practice.

Nancarrow et al. (2013) combined a large survey with an extensive literature review to identify the attributes of a good interdisciplinary team using qualitative content analysis. They identified 10 themes that support effective interdisciplinary teamwork. We have adopted eight of these: leadership and management; communication; appropriate resources and procedures; appropriate skill mix; climate; individual characteristics; clarity of vision; and respecting and understanding roles. The other two, personal rewards, training, and development; and quality and outcomes of care, are not relevant to our CI.

Many studies have identified competencies relating to teamwork and collaboration. Communication within the team (Molyneux, 2001), team structure (Xyriches & Lowton, 2008), and knowledge integration (Claus & Wiese, 2019) were the most frequently occurring in our review, which provided excellent guidance for our work with the ACI design team.

Stages of Design

In this section, we break down the creation of the ACI into five distinct stages, detailing the formative decisions and obstacles encountered during the design process.

Recruiting the Design Team

Algebra plays a central role in second level mathematics education and continues into a variety of pure and service mathematics undergraduate courses. The wide range of uses of algebra, coupled with the different applications each course may focus on, can require the emphasis of different facets of the content area. Given this, it was important that the ACI design team included members from as many different stakeholders as possible. The expertise of the following groups was acknowledged: Lecturers in Mathematics, Engineering, Physics, Business, other service subjects, and Teacher Education; Members of the Mathematics, Science, and General Education communities; Members of the Mathematics Support community; Second level Mathematics teachers; and Students.

While a representative from each stakeholder group would be optimal, logistic reasons would make it impractical. This is in line with the thoughts of Xyriches and Lowton (2008) that team size and composition are key themes in interprofessional teamwork. It was decided that a team of seven members[1] would begin the project. Each of the team members possess expertise relating to more than one of the above roles, but members of the general education community, a second level mathematics teacher, and a student voice were omitted from the

[1] Dr Eabhnat Ní Fhloinn (DCU) and Dr Micheal Carr (TU Dublin) were recruited to the design team and attended the first meeting of the team where significant progress was made on defining the parameter of the ACI. Their involvement beyond this stage was curtailed by workload and we would like to thank them for their work on the project and their continued support.

design team. This decision was made to keep the team at a manageable size and with the knowledge that their perspectives would be heard and incorporated later in the process.

Defining the Parameters of the ACI

Once the design team was formed, the priority was to establish the meaning and intention of the ACI before advancing. Algebra is an extensive content area within mathematics that begins at primary level and extends beyond undergraduate level. This can cause uncertainty with respect to the facets of algebra that should be included in the inventory. To ensure that the design team and future administrators understood the scope of the ACI, the following definition of this scope was agreed: *The elements of algebra a student should understand having completed second level education.* There were several factors that influenced our definition, most notably previous work on CIs and an algebra decomposition (Figure 1), discussed below. It is common for CIs to discuss what their threshold for inclusion is, for example, Epstein (2013, p.1018) discussed in relation to the Calculus Concept Inventory (CCI) "… a set of very basic concepts that all sides agree students should – must – be expected to master in, for example, first semester calculus". The team agreed that the wording used by Epstein provided clarity (for the CCI) and a scaffold from which we could work. Though Epstein elected to use the first semester of tertiary level as a time stamp, we preferred the end of second level. At this stage in their education, students will have studied the same mathematics syllabi (at primary and second level) albeit to different standards. All higher and further education programmes will have different approaches to mathematics that will lead to an increasingly heterogeneous population as they continue through their education. Second, including this time stamp also benefitted our algebra decomposition because we could use the senior cycle syllabus (NCCA, 2021) to inform our work.

The idea of building a decomposition for the subject in focus is common to CI studies. It began with Hestenes et al. (1992, p.142), who developed a table that separated the force concept into six 'conceptual dimensions', all of which are necessary for a complete conception of force. Other CI designers have proceeded in a similar way (e.g., O'Shea et al. 2016;). Our decomposition was developed through multiple meetings of the design team. Initially, each member shared a preliminary decomposition with the group to form an exhaustive list of content for inclusion. This was examined by a team member, who removed redundant entries and gave structure to the decomposition. This version of the decomposition was returned to team members who made further suggestions on content and structure. Multiple rounds of feedback and revision took place until the group reached a consensus about the decomposition. Figure 1 details our deconstruction of the concept of algebra.

Figure 1

Algebra Decomposition

Equality	Solution	MER
Meaning of '=' • Evaluate vs equivalent to • Balance model • Symmetry • Transitivity • Misuse (≠, =>) • Identity **Equations** • Rules for solving equations • Properties of operations (commutativity, associativity, etc.)[2] **Inequalities** • Rules for solving inequalities	**Meaning of solution(s)** • Equations vs inequality **MER** • Algebraic • Graphical (x-intercept, POI,) **What's valid vs invalid**	**Fractions as:** • Parts of whole • Size of portion • Quotient • Ratio • Decimal • Percentages • Graph/chart **Word problems to equations and vice versa**
Expressions	**Variables**	**Operations**
Fractions **Polynomials of order ≤2**	**Variable as:** • Constant • Placeholder • Multiple values simultaneously **Zero** **Proportional reasoning**	**Manipulations with all operations** • +, -, x, ÷, (), exp., log. **Expanding, factorizing, transposing, simplifying** **PEMDAS**

Our decomposition divides algebra into six sections: Equality, Expressions, Solution, Variables, Multiple External Representations (MER)[2], and Operations. Each comprises smaller subsections which detail the intricate aspects of each section that is included in our decomposition. Some subsections apply to more than one section (e.g., Fractions, MER) and appear as often as required for the reader to interpret our algebra decomposition. In this sense, we consider the decomposition to be exhaustive, with the exception of properties of operations. Though they are in the decomposition under Equality, including each property for each operation would be impractical. Ultimately, the team decided to acknowledge their importance through inclusion in the decomposition, consider items assessing them in our shortlist, and then prioritize the most important ones during the creation of the ACI.

Shortlisting Potential Items

Shortlisting questions for the ACI involved gathering many items that, collectively, assess each aspect of the decomposition. The algebra decomposition provided tremendous clarity to the task and was a constant point of reference that aided communication and fast-tracked many aspects of this phase of the project. Using an online space for collecting and developing items was necessitated by COVID restrictions but was also helpful when refining tasks and is recommended by Carlson et al. (2010).

The shortlisted items came from multiple sources. Research articles, algebra tests, and textbooks were consulted, and several items were adapted for the ACI. Algebra tests included a sample of international assessments (e.g., TIMSS), national assessments, and mathematics

[2]We use MER (e.g., text, pictures, equations) in the same manner as Ainsworth (1999), who explored the different ways MERs can be used to support learning and detailed a functional taxonomy of multiple representations. An example of multiple representations in mathematics would be the use of graphs, tables, and equations when teaching functions.

diagnostic tests at third level. These are indicative of what is valued by assessors and provide a rich pool of items. Previous research on algebra and task design were also studied for items, and a significant number of items were developed by members of the design team. All methods contributed to our shortlist, which began with over 100 items that mapped to at least one facet of the decomposition. Team-developed items were more frequent, primarily in responses to specific facets of the decomposition that are less common in testing or textbooks.

Once collected, each team member provided feedback on all items. The feedback was collated and reconciled which resulted in a refined list of fewer items. This list was returned to each team member for further feedback, beginning an iterative process which occurred several times. Sometimes new items were added as they were discovered, or items were created in response to discussions around shortlisted items. As each round of revisions occurred attention to detail increased, and the standard for inclusion became higher, with later rounds being stricter wording, ordering, and general presentation, in addition to the mathematical merit of each item. Factors considered at this stage are the students' familiarity with notation, and ability to interpret the question as intended. Wage et al. (2005) highlighted issues such as how the questions are presented to students (linked to MER) and whether questions focused a single concept from the decomposition or required students to use knowledge of multiple concepts. Similarly, Halloun and Hestenes (1985) were forced to remove two problems from their assessment because, despite being well-posed, they were misinterpreted by students during interview more often than not.

Each question was developed and administered as a multiple-choice question (MCQ). We chose MCQs because they are simpler to complete and to analyze than open response items. Halloun and Hestenes (1985, p. 1044) concluded that the MCQ version of their assessment 'measures the same thing as the written version but more efficiently'. Interviews with students will be conducted later in the project. This will inform distractor – an incorrect option on an MCQ – choice, and even item inclusion. We planned to administer a written version of the ACI to students to select distractors, as done by Halloun and Hestenes (1985) and others. However, lack of access to students made this impossible. We chose three options per question for the ACI because we were keen to minimize the word count of the ACI as much as possible. Vyas and Supe (2008) include reading time among the numerous practical advantages of including three options over four or five. More importantly, they claim that there was "no significant change in the psychometric properties of the 3 options test when compared with 4 and 5 options" (p.130).

Another motivation for reducing the word count of the ACI was to accommodate a Certainty of Response Index (CRI) for each question. The CRI method is used to distinguish between lack of knowledge and misconceptions at both individual and group level. The approach was used by Hasan et al. (1999) on Halloun & Hestenes' (1985) Mechanics Diagnostic Test. In essence, the authors (Hasan et al., 1999) used a CRI with each question to associate how confident the student was in their response. Responding with low confidence (independent of the student's correctness), indicates a lack of knowledge. High confidence with a correct answer is a justification of the student's confidence in their answer, but high confidence attached to an incorrect answer indicates a misconception. In total, over 100 items were reduced to 31 using the iterative design process.

Piloting the ACI

Piloting the ACI began as soon as the design team had agreed on the items to include. This allowed individuals outside of the design team to offer insights beyond those of the design team. Piloting the ACI also allowed for the online format of the ACI to be tested before rollout began[3].

The piloting and refinement of CIs is well described in the literature. Hestenes et al. (1992) recruited lecturers and students to provide feedback on their tests prior to rollout. O'Shea et al. (2016) included second level teachers in their study. Interviewing students who have taken the CI in question (often called cognitive laboratories) is also commonly done (Hestenes et al., 1992) to learn more about the answers provided and associated reasoning..

During piloting, the ACI was shared with representatives of the following groups: a lecturer in Mathematics and Teacher Education, a member of the Mathematics Support community, a second level Mathematics Teacher, and a student. We were particularly interested in their opinion on the language and symbols used. As outlined above, well-posed questions can often be read in an unintended manner by students or may contain terminology that is beyond the level of the test. The feedback we received during piloting was shared with the design team which led to minor changes (all of which pertained to the foreword preceding the ACI). This resulted in a 31-item preliminary ACI, advertised to students in February 2021.

Data Collection and Validation

The collection of data is a vital stage in the validation of a CI. Questions of reliability and validity are significant for CIs and require many participants to engage with each item. The FCI's reliability was based primarily on the Mechanics Diagnostic Test (MDT), a precursor to the FCI, from which over half of its items are taken. Halloun and Hestenes (1985) used four mechanisms 'to establish the face and content validity' of the MDT: feedback from experienced lecturers; testing of graduate students; interviewing undergraduate students; in-depth analysis of high achievers' answers.

The most detailed analysis outside of physics was carried out by Steif and Dantzler (2005) on their Statics CI. They ascertain their CI's reliability, content validity, criterion-related validity, and construct validity through numerous calculations (Cronbach's alpha, Spearman's rho, confirmatory factor analysis, etc.). O'Shea et al. (2016) used Rasch Analysis, as well as Cronbach's alpha, to investigate the validity and reliability of their test. Carlson et al. (2010) discuss internal and external content validity and report going through an iterative process of administering the Precalculus Assessment (PCA) and carrying out follow-up

[3] CIs are traditionally pen and paper tests. Administering a pen and paper test currently is not possible and so the decision was made to convert the ACI into an online exam.

interviews with students. They also 'examined external measures to further establish PCA's validity as a tool for determining students' preparedness for beginning calculus … by correlating students' post-course PCA scores with their course grades (p.124)'.

Data collection relies on the success of its rollout strategy. The ACI is no exception, and the effort made by the extended mathematics education community and other departmental colleagues was of paramount importance, especially considering that in-person testing was not possible. Our rollout strategy comprised two parts: personal emails to lecturers with whom we had a previous working relationship, and general emails sent to mailing lists. The ACI went live in February 2021 and received 330 responses as of April 2021.

Discussion

In this paper, we detail the design of an Algebra Concept Inventory (ACI) by an interdisciplinary research team. We found Nancarrow et al.'s (2013) 10 themes to support interdisciplinary teamwork very useful to our project. Within this, team structure, communication, and knowledge integration were the most applicable.

The results will be used to design teaching materials that will be used with subsequent cohorts of students nationwide. The initial data from the ACI has already highlighted specific aspects of algebra about which students possess misconceptions. These misconceptions certainly occur in all HEIs, and we intend to make our resources freely available. We hope to continue to increase the number of responses to validate our ACI instrument.

A validated CI can be used as a diagnostic tool, for evaluating instruction, and placement exams (Hestenes et al., 1992) which offer exciting avenues to extend the research. The rapid, pandemic-induced transition to online teaching in HEIs has accelerated work that focuses on the affordances of online and blended instruction (Hyland & O'Shea, 2021). One avenue to pursue is that of a personalized, adaptive approach to teaching and learning (Walkington, 2013). The ACI could have a role in 'triaging' students for entry onto such a flexible learning pathway, and for determining appropriate progression. Hestenes et al. (1992) talk about threshold scores in the FCI near 60% and 80%. Failure to achieve 60% means "a student's grasp of Newtonian concepts is insufficient for effective problem solving" whereas achieving 80% is indicative of a true Newtonian thinker. Corresponding thresholds could be identified in relation to the ACI to inform decisions about entry points and pathways.

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33*(2-3), 131-152.

Carlson, M., Oehrtman, M., & Engelke, N. (2010). The precalculus concept assessment: A tool for assessing students' reasoning abilities and understandings. *Cognition and Instruction, 28*(2), 113-145.

Claus, A., & Wiese, B. (2019). Development and test of a model of interdisciplinary competencies. *European Journal of Work and Organizational Psychology, 28*(2), 191-205.

Epstein, J. (2013). The calculus concept inventory-measurement of the effect of teaching methodology in mathematics. *Notices of the American Mathematical Society, 60*(8), 1018-1027.

Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics, 53*(11), 1043-1055.

Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the certainty of response index (CRI). *Physics Education, 34*(5), 294-299.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141-158.

Hyland, D. & O'Shea, A. (2021). *The student perspective on COVID-19 related closures at Irish universities.* In review.

Molyneux, J. (2001). Interprofessional teamworking: what makes teams work well?. *Journal of Interprofessional Care, 15*(1), 29-35.

Morse, W. C., Nielsen-Pincus, M., Force, J. E., & Wulfhorst, J. D. (2007). Bridges and barriers to developing and conducting interdisciplinary graduate-student team research. *Ecology and Society, 12*(2).

Nancarrow, S., Booth, A., Ariss, S., Smith, T., Enderby, P., & Roots, A. (2013). Ten principles of good interdisciplinary teamwork. *Human Resources for Health, 11*(1), 1-11.

National Council for Curriculum and Assessment. (2021). Curriculum online - Mathematics [online]. Available from: https://www.curriculumonline.ie/Senior-cycle/Senior-Cycle-Subjects/Mathematics/ [accessed 1 April 2021]

O'Shea, A., Breen, S., & Jaworski, B. (2016). The development of a function concept inventory. *International Journal of Research in Undergraduate Mathematics Education, 2*(3), 279-296.

Steif, P. S., & Dantzler, J. A. (2005). A statics concept inventory: Development and psychometric analysis. *Journal of Engineering Education, 94*(4), 363-371.

Stone, A., Allen, K., Rhoads, T. R., Murphy, T. J., Shehab, R.L., & Saha, C. (2003). The statistics concept inventory: A pilot study. In *33rd Annual Frontiers in Education, 2003. FiE 2003.* (Vol. 1, p. T3D-1). IEEE

Vyas, R., & Supe, A. (2008). Multiple choice questions: a literature review on the optimal number of options. *National Medical Journal of India, 21*(3), 130-3.

Wage, K. E., Buck, J. R., Wright, C. H., & Welch, T. B. (2005). The signals and systems concept inventory. *IEEE Transactions on Education, 48*(3), 448-461.

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology, 105*(4), 932.

Xyrichis, A., & Lowton, K. (2008). What fosters or prevents interprofessional teamworking in primary and community care? A literature review. *International Journal of Nursing Studies, 45*(1), 140-153.