

Building epistemic thinking through disciplinary inquiry: Contrasting lessons from history and biology

Michael Johnston, Rosemary Hipkins, and Mark Sheehan

<https://doi.org/10.18296/cm.0020>

Abstract

This article explores the effect of high-stakes assessment on the representation of epistemic knowledge in the enacted curriculum—that is, the curriculum experienced by students in the classroom. Epistemic knowledge concerns the processes for constructing and evaluating theories that explain phenomena in the natural and social worlds. Knowledge-building disciplines such as history and science each have their own epistemic processes. We explore the extent to which these processes are reflected in the standards used to assess history and biology for the National Certificates of Educational Achievement (NCEA). We show that these processes are not well represented in the externally assessed (examination-based) standards for either discipline, and that biological epistemology is not well represented by its internally assessed standards either. The internally assessed standards for history, however, do involve students in a simplified version of authentic historical enquiry. In a statistical component of the research, we show that internally assessed standards for history are a stronger predictor of subsequent achievement in history than the externally assessed standards for history, whereas the converse is the case for biology. We suggest that the epistemic focus of the internally assessed standards in history has resulted in the enacted curriculum for this subject being more epistemically based than is the case for biology.

Introduction: What is epistemic thinking and why does it matter?

This article reports on an investigation of the kinds of knowledge that are valued in the achievement standards used to assess senior secondary

students' learning in the disciplines of history and biology. This is an important question for the senior school curriculum in these disciplines because of the way in which high-stakes assessment influences what is taught and learnt (for example, Rust, 2002).

More specifically, we examine the extent to which the history and biology achievement standards promote opportunities for students to demonstrate epistemic thinking. We also investigate the extent to which such opportunities are predictive of progress in these disciplines. We conjecture that students' understanding of the epistemic foundations of a discipline are likely to be predictive of their progress in that discipline and that, irrespective of the status of such knowledge in the *published* curriculum, if epistemic thinking is promoted by high-stakes assessment, then it is more likely to feature in the *enacted* curriculum, which is the curriculum actually experienced by students.

We broadly differentiate three types of knowledge in relation to the learning of disciplines such as history and biology:

- **Declarative** knowledge is propositional in nature. An emphasis on declarative knowledge in a curriculum promotes an approach to learning based on memorisation and recall.
- **Conceptual** knowledge organises otherwise disparate information into coherent systems. An emphasis on conceptual knowledge encourages students to learn to link ideas and to understand past and present theories within the explanatory domain of a discipline.
- **Epistemic** knowledge relates to the disciplinary inquiry processes used to construct and test new theories. An emphasis on epistemic knowledge promotes understanding of the process of knowledge production in a discipline; for example, how we know what we know, and on what basis certain knowledge claims might be judged more valid than others.

Declarative knowledge provides an essential foundation for building conceptual links, and both declarative and conceptual knowledge underpin epistemic thinking; that is, the application of epistemic knowledge to specific questions. Thus, epistemic knowledge is defined by an *understanding* of knowledge-production methodologies and epistemic thinking is defined by the *application* of such methodologies. The latter

provides unique opportunities for students to build important capabilities that they will not necessarily develop without specific pedagogical support. Illustrating this need, a recent futures-thinking exercise by Hipkins, Bolstad, Boyd, and McDowall (2014) worked backwards from a small selection of complex contemporary problems to identify the capabilities students need for critical citizenship. For example, one feature of complex multifaceted problems is clearly illustrated by a consideration of claims made in relation to climate change: different groups make conflicting claims about the underlying basis of the phenomenon, and therefore about what should be done in response. Determining who to trust, and on what basis, is a necessary part of becoming informed and deciding how to act. In a complex world in which the future is largely unknowable and appears to be changing with ever-increasing rapidity, it is not enough for students to be receivers of declarative and conceptual curriculum knowledge without also learning about the processes by which it is generated in different disciplines, and more importantly still, how to apply these processes to novel situations (Barnett, 2004).

Developing epistemic thinking capabilities is also critically important for teachers. In New Zealand, teachers are confronted with endless potential choices for building a senior secondary school curriculum based on a framework curriculum document, *The New Zealand Curriculum (NZC)* (Ministry of Education, 2007) and a modular assessment system, National Certificates of Educational Achievement (NCEA). In theory, the latter is aligned with the former, but in practice the alignment is sometimes superficial or only partially realised. Having a secure epistemic understanding of at least one discipline area can help teachers navigate important pedagogical choices, including: building coherent courses and overall programmes of study (Hipkins, Johnston, & Sheehan, 2016); maintaining the integrity of individual discipline areas when creating integrated (cross-curricular) courses of learning (for example, Colucci-Gray, Trowsdale, Davies, Burnard, & Gray, 2017); and guiding students as they carry out participatory and increasingly open-ended inquiries of their own (for example, Osborne, 2014; Tytler, 2016). These challenges are not unique to New Zealand, but they are certainly exacerbated by the great freedom for individual and school-level curriculum making in our education system (Zohar & Hipkins, 2017).

Epistemic thinking as a recent curriculum and assessment focus

The most recent Programme for International Student Assessment (PISA) framework states that students' epistemic knowledge will be assessed as an important aspect of scientific literacy. For students at age 15 (that is, just before their senior secondary years) the ideas in scope for demonstrations of epistemic thinking are outlined as follows:

Epistemic knowledge includes an understanding of the function that questions, observations, theories, hypotheses, models and arguments play in science; a recognition of the variety of forms of scientific enquiry; and the role peer review plays in establishing knowledge that can be trusted. (Organization for Economic Cooperation and Development [OECD], 2016, p. 19)

The science learning area of *NZC* has an overarching Nature of Science (NOS) strand that should address elements of epistemic knowledge such as those outlined in the PISA definition. There is also a clear signal in the “essence statements” in *NZC* (p. 17) that science learning is first and foremost important to develop capabilities for citizenship, which would include epistemic thinking (see Bunting, Cowie et al., 2017). However, whether or not students develop epistemic awareness and thinking depends on how teachers actually understand the purpose of this NOS strand. They might simply ignore it, or add an extra set of declarative propositions to the curriculum they already teach (Hipkins, 2012).

There is also a signal in *NZC* that epistemic thinking is important in the subject of history. The achievement objectives at Levels 6–8 direct attention to ideas such as “cause and consequence” and “significance [of events and of evidence]” (*NZC*, fold-out pages). These ideas are central theoretical constructs that guide the knowledge-building work of historians, and their importance in the school curriculum is signalled by the use of the term *historical thinking* (Seixas & Morton, 2012; Shemilt, 1983). However, just as science teachers might neglect signals given by the NOS strand of *NZC*, there is evidence that history teachers often make narrow curriculum choices that don't provide their students with the broad, in-depth understandings of the past that are a core feature of critical citizenship (Ormond, 2017).

Teachers have reported that their (enacted) curriculum choices in the senior secondary school years tend to be dominated by NCEA rather than directly by *NZC*. This holds right across the curriculum learning areas and has been the case since the inception of NCEA (Hipkins, 2013). Indeed, this challenge was not instigated by NCEA: high-stakes assessment dominates teachers' pedagogical decision making much more generally (see Rust, 2002). However, because NCEA now provides the framework for senior school assessment, it is the focus of the various achievement standards that influence the manner in which teachers shape the curriculum that students experience—which, as we have already noted, may contrast starkly with the published curriculum.

In relation to *NZC*–NCEA dynamics, the subject of history provides an interesting contrast to the sciences, given our focus on epistemic thinking: Since the *NZC*–NCEA alignment process, carried out between 2011 and 2013, the achievement standards for history have reflected more explicitly the epistemic features of the parent discipline, as signalled in *NZC*, than is the case for sciences. Compared with history, references to epistemic knowledge and thinking in the science achievement standards tend to be fewer and more token in nature (Hipkins, 2012).

NCEA “inquiry” as an expression of what is valued in the curriculum

Our study focuses on the NCEA achievement standards in history and biology that are concerned with inquiry in these disciplines. We chose these two subjects because the epistemic processes of their parent disciplines provide interesting contrasts and, as noted above, some achievement standards in history provide more explicit epistemic signals than those in the sciences. In a qualitative analysis we focused on the content and approach to assessment in the inquiry achievement standards in each discipline. Our reasoning was that inquiry is a core aspect of any epistemic process, and the manner in which students (and teachers) experience and come to understand inquiry will therefore shape their epistemic thinking—even if they are not aware of this. All the inquiry standards are internally assessed by teachers.

Our analysis of the inquiry standards was informed by a recent *practice turn* in inquiry pedagogy. This change in emphasis is evident in curriculum-related materials elsewhere; for example, in the common core science standards in the United States (Osborne, 2014). Osborne draws attention to the powerful potential impact of the change: rather than focusing on pre-formulated “inquiries” that lead to outcomes already known by the teacher (as in traditional school science experiments or common types of topic-based inquiries in other subjects) students are supported to directly experience the epistemic practices of knowledge building within a discipline. In short, they carry out more authentic inquiries in which the findings are not necessarily predetermined. One American educator uses the metaphor of supporting students to play “junior versions of the whole game” of knowledge-building (Perkins, 2009). This metaphor—playing simplified forms of adult sports—helps to demonstrate how students might learn to inquire into issues of authentic concern to them and to their communities, in ways that accord with expert practice, albeit supported, so as to be manageable and meaningful at their current stage of schooling.

The practice turn deliberately aims to foster epistemic awareness by engaging students in the processes of theory construction, evidence collection, and evidence evaluation, as well as learning the more formal structure of a discipline. It might be claimed that traditional inquiry practices already do this, but over the last several decades important theoretical insights have arisen that would challenge such a claim. Anthropological studies of actual disciplinary practices have shown that there are differences between the formal conception of a discipline, as might be conveyed by more traditional curriculum materials, and its practice. For example, in experimental science, the formal process of testing a theory might require an attempt to experimentally falsify one or more hypotheses distilled from that theory. In practice, however, scientists often begin with less formal observations to establish the viability of a theoretical idea before engaging in the more formal processes of systematically gathering and analysing empirical data. That is not to say that the formal process is unimportant—indeed it is vital—but if a discipline is taught only as an abstract, formal process, there is a risk that students will misunderstand the ways in which disciplinary researchers actually operate. Thus, to be effective, epistemic pedagogy

must involve students in active authentic inquiry—via “junior versions” of what practitioners in the relevant disciplines actually do.

The practice turn implies a further pedagogical and assessment challenge, relating to the conceptualisation of “progress”. Drawing on sociocultural frameworks, Tytler (2016) proposes three distinct, contextually rich dimensions, which begin to describe a notion of progress within an epistemically focused, participatory curriculum:

- Students show that they can deploy an expanding and increasingly nuanced repertoire of inquiry practices relevant to the discipline (i.e., the practice turn in action).
- Students learn to see beyond local causal processes to recognition of large-scale systems and patterns.
- In participatory terms, students show an ability to go beyond the local epistemic practices of the classroom to more active participation in wider cultural landscapes. This includes interacting with professionals from relevant discipline areas.

Tytler argues that this view of progress encompasses increasing complexity and refinement of: the substantive and syntactical structures of the discipline; the representation and reasoning processes of the discipline; and the “meaningful production of knowledge in expanding cultural landscapes” (p. 6). All these have a distinctly epistemic focus overlaid on the knowledge and skills that constitute traditional curriculum “content”. If students are indeed making this type of progress, we could anticipate that their ability to demonstrate rich conceptual knowledge will be strengthened along with their ability to work with the inquiry processes of the relevant discipline. In other words, we might predict that overall achievement will be lifted if and when students have powerful experiences of authentic inquiry.

Methodological approach

The inquiry reported here has two parts, which combine to test the hypothesis that educational achievement in a disciplinary framework is positively influenced by building epistemic thinking through authentic experiences of disciplinary inquiry.

Building epistemic thinking through disciplinary inquiry

In a qualitative component we explored the kinds of activities that students are likely to engage in when completing the internally assessed achievement standards involving disciplinary inquiry. In particular, we investigated the extent to which the standards and associated materials afford students opportunities to demonstrate epistemic thinking and, concomitant to the influence of assessment on the enacted curriculum, the extent to which these standards and materials are likely to foster the teaching and learning of epistemic thinking.

We conjecture that internal assessment, including—but not limited to—the inquiry standards themselves, provides a more feasible basis than external assessment for the assessment of epistemic thinking. This is because external assessment in biology and history comprises entirely time-limited examinations, which afford neither the time nor the scope to engage in anything approaching authentic epistemic processes. Internal assessments on the other hand, offer opportunities for extended projects that can be conducted in a variety of venues and to apply epistemic thinking to questions especially chosen to be of interest to particular groups of students. To explicitly compare the two assessment formats in this regard, we extended our qualitative analysis to selected externally assessed standards to examine the types of knowledge (declarative, conceptual, or epistemic) that are needed to satisfy the criteria for achieving these standards.

A statistical component of the inquiry complemented the qualitative component by comparing biology and history for the extent to which the internal and external assessment formats of each discipline best predict attainment in each assessment format in the following year. The logic of this analysis is that, if epistemic thinking is well represented in assessment for a discipline—whether internal, external, or both—then, because of the influence of assessment on the enacted curriculum, that assessment will promote the teaching and learning of epistemic thinking.

To the extent that engaging in the knowledge-building (epistemic) processes of a discipline fosters a coherent conceptualisation of that discipline, the amount of experience students have with such processes is likely to be correlated with their achievement and progress. Thus, if our qualitative analysis shows that internal assessment provides stronger opportunities than external assessment for developing epistemic thinking

then, provided that these opportunities are appropriately exploited, we would expect the statistical analysis to show that achievement in internal assessment would predict achievement in the same discipline in the following year more strongly than external assessment. That any epistemic opportunities are indeed appropriately exploited is an important qualification, especially because NCEA allows teachers a great deal of flexibility in respect of the assessment process for any given standard. Confirmation that any such predictive relationship is attributable to an epistemic focus in internal assessment would therefore require closer observation of the relationship between the representation of epistemic thinking in a range of specific assessment activities and the subsequent progress of students undertaking each activity.

NCEA inquiry in biology and history

Practical inquiry in biological contexts is represented by internally assessed achievement standards at all three NCEA levels. Given that students are required to gather empirical data in some form, these inquiries should arguably afford an opportunity for students to participate in a simplified form of the “whole game” of knowledge-building in biology. However, the achievement criteria for these standards, described in Table 1, provide only weak signals that epistemic thinking should be an important learning and assessment focus. Notice that the sense of progress conveyed by the titles of the standards rests on the level of support provided by the teacher rather than on any increase in the sophistication of inquiry practices of the sort described by Tytler (2016). Particularly at Level 1, there is a formulaic structure, and the “cultural landscape” of the inquiry activity is firmly centred in the school laboratory.

A comparison of the expectations of *NZC* with the criteria for these achievement standards does little to increase confidence that any impetus in *NZC* for a shift to an epistemic focus is supported by NCEA. According to *NZC*, at Level 6—which is expected to align with NCEA Level 1—students should develop and carry out more complex investigations, involving the use of models and multiple variables. However, the achievement standards do not include model-based investigations as a suggested option until Level 2.¹ The formulaic investigation described

Building epistemic thinking through disciplinary inquiry

Table 1. A brief overview of *investigation* achievement standards in biology

Number/ name/ level	Brief commentary on assessment focus/ how achievement is differentiated (A/M/E)
90925: Carry out a practical investigation in a biological context, with direction (L1)	<p>Students are given a pre-determined question to investigate. They convert the question to a hypothesis and 'design' a method to address it (either a fair test or a pattern seeking method is expected). A suggested template is provided for the written report that will be assessed.</p> <p>Students gain Merit if they also include a discussion of the biological ideas relevant to their investigation. They gain Excellence if they add an element of critique such as justification of the chosen method, evaluation of the reliability of the data or etc., or if they demonstrate deeper and more comprehensive knowledge of the biological ideas relevant to the investigation.</p>
91153: Carry out a practical investigation in a biological context, with supervision (L2)	<p>Students are given an investigative context but have somewhat more freedom than at Level 1 to choose their own questions and to design the associated inquiries. Field work could well be involved and more learning time might be provided than at Level 1. There is no pre-specified report template. Support might be provided in the form of whole-class discussion before students write their individual formal reports.</p> <p>Merit and Excellence are determined by additional detail similar to that described for the level 1 standard.</p>
91601: Carry out a practical investigation in a biological context, with guidance (L3)	<p>A more "open-ended" investigation is anticipated and students can be given opportunities to make changes to their initial designs once their investigations are underway. They keep a logbook to record their investigative thinking and activities but this is primarily used to ensure authenticity. Again, the formal report is what is marked. Students are expected to use secondary sources of data in addition to the primary data generated by their own investigations.</p> <p>Again, Merit and Excellence are determined by additional detail similar to that described for the level 1 standard.</p>

for Level 1 could hardly be described as “complex” and neither will it necessarily afford an opportunity to explore the challenges raised by multiple variables.

The published exemplars of the criteria for each grade² for the Level 1 inquiry standard support the initial impression that students need to produce a traditional laboratory-based investigation to confirm an already well-known biological relationship, using familiar “fair testing” experimental protocols. The predetermined nature of the anticipated outcomes stands in contrast to the open-ended nature of more authentic

inquiries, as outlined above. Furthermore, the detailed notes provided with the exemplar materials strongly imply that the main focus should be on how well students have understood the declarative knowledge of the curriculum, rather than on any demonstrations of epistemic thinking. Exemplars for the Levels 2 and 3 biological inquiry standards have a similar emphasis to that just outlined: work deemed “high quality” tends to demonstrate deeper knowledge of theoretical rather than epistemic constructs.

Table 2 shows an analysis of the criteria for the inquiry-based standards in history, which are analogous to the investigation standards in biology, in that they relate more closely than other standards to the epistemological processes of the discipline. They contrast strongly with the biology standards in the extent to which they enable students to experience a simplified but authentic version of the process of historical inquiry. The contribution that internally assessed coursework in history makes towards motivating young people to develop epistemic understandings of the past (including the capacity to think independently and adjudicate between competing claims of historical authenticity) was apparent in a 2011–13 TLRI study (Sheehan, 2013a,b). In large part this is because students were involved in inquiry processes that emulated how historians (as experts in the domain) generate and evaluate knowledge.

The *NZC* achievement objectives require students to demonstrate a grasp of the research protocols of the discipline of history and to, some extent, this is reflected in the achievement standards at all three levels of NCEA. Unlike biology, in which the sense of progression is largely given by a successive reduction in the degree of supervision by teachers, the history standards demonstrate a trajectory describing a progression of historical (epistemic) thinking. While the differentiation between Achieved and Merit grades is not always clear, the exemplars demonstrate a clear epistemic progression between Achieved and Excellence within each NCEA level, as well as between the criteria for attaining credit across year levels. It is not simply a matter of describing historical phenomena in successively greater detail. In this regard, the standards themselves, and the exemplified criteria, indicate that these assessments reflect the epistemic intentions of *NZC* in the history learning area.

Table 2. A brief overview of investigation standards in history

Number/ name/ level	Brief commentary on how achievement is differentiated (A/M/E)
<p>91001: Carry out an investigation of a historical event, or place, of significance to New Zealanders (L1)</p>	<p>There are no specified questions or contexts. Approaches to choosing questions and topics range from being entirely student-driven to being teacher-directed, using focusing questions. Year 11 (NCEA Level 1) is the first year in which history is taught as a discipline-informed subject, so teachers typically provide substantial support for students engaged in research.</p> <p>Students gain an Achieved grade if they are able to identify historical sources of evidence, record these appropriately, and establish why the sources are relevant and significant. To gain Merit they need to demonstrate that they have gathered evidence from a variety of historical sources and evaluated their investigative process. Excellence requires them to establish links between evidence and questions and to evaluate the usefulness of the evidence.</p>
<p>91229: Carry out an inquiry of a historical event or place that is of significance to New Zealanders. (L2)</p>	<p>As at Level 1, there are no specified questions or contexts and while some teacher support is typically provided at early stages, students work largely independently.</p> <p>Students gain Achieved if they can demonstrate that they are able to carry out and evaluate an inquiry and to annotate and organize sources of evidence. Merit places more emphasis on evaluating the strengths, weaknesses, successes and difficulties of the inquiry process as well demonstrating higher-order thinking skills. To gain Excellence, candidates are additionally required to explain the comparative usefulness and reliability of sources of evidence.</p>
<p>91434: Research a historical event of place of significance to New Zealanders, using primary and secondary sources (L3).</p>	<p>Again, there are no specified questions or contexts. Students have a high degree of choice over the focus of their course work and are expected to work independently. Students gain Achieved if they are able to evaluate the research process, comment on the relevance of selected evidence and on the reliability and usefulness of sources. Merit is not well distinguished from Achieved: Candidates are still required to evaluate the research process, comment on the relevance of selected evidence, on the reliability and usefulness of sources. The difference is that, for Merit, these comments should be in-depth; however, what in-depth means is not well defined. It is clearer how Excellence is differentiated; candidates are required to show initiative in gathering and selecting evidence and to analyze, as well as evaluate, the research process, including the reliability and usefulness of specific sources.</p>

The emphasis of external assessments in biology and history

Like most NCEA external assessments, all the external biology and history achievement standards are assessed under formal, time-limited examination conditions. Table 3 compares three of these standards in biology, chosen to represent as closely as possible expected progress in the “Life processes” sub-strand of *NZC* at Levels 6–8. The clear pattern is of a requirement for a traditional demonstration of declarative and conceptual knowledge. Students are expected to recall and discuss progressively more difficult concepts at higher NCEA levels. Within each level Merit is demonstrated by being able to explain the concepts in addition to simply recalling them, and Excellence, by deploying higher order thinking skills in order to discuss some less-straightforward application of the knowledge being assessed. At no point, however, is any requirement for epistemic thinking evident.

As is the case for the internally assessed standards, there are no specified contexts for the externally assessed history standards, although there are suggestions that candidates and teachers may follow, and teachers typically provide guidelines and follow common programmes that prepare students for the examinations.

Table 4 compares three externally assessed history standards that assess *cause and consequence*, which is a key epistemic concept in history. In these standards, candidates are required to demonstrate a grasp of how the historical thinking concept of cause and consequence (as set out in *NZC*) frames our understanding of the past. At Level 1 NCEA (Level 6 *NZC*) the learning area prioritises “causes and consequences of past events that are of significance to New Zealanders” and different “perspectives” on the past. At Level 2 (Level 7 *NZC*) the emphasis is on “interpretations” of the past and how “historical forces and movements have influenced the causes and consequences”. However, while this epistemic progression is apparent in the achievement standard documents themselves, it is not apparent in the exemplars. The differentiation between Achieved and Merit is far from clear and it appears to be largely a matter of gathering more detail. This is especially apparent in the judgements made between

Table 3. Summary of three externally assessed biology achievement standards

Number/ name/ level	Brief commentary on assessment focus/ how achievement is differentiated (A/M/E)
90927: Demonstrate understanding of biological ideas relating to micro-organisms (L1)	The notes for this standard prescribe the types of biological ideas that might be assessed in the examination. These notes include a list of specific ideas with which students must be familiar: “inoculate, extracellular digestion, enzyme, parasite, pathogen, toxin, decomposer, aerobic, anaerobic, hyphae, binary fission, spores, sporangium, antibiotic and vaccine”. Merit is differentiated as <i>explaining</i> these ideas and Excellence by <i>linking related concepts together</i> in an elaboration or justification.
91156: Demonstrate understanding of life processes of plants and animals at the cellular level (L2)	Three specific life processes are given in the range notes: photosynthesis, respiration and cell division. Students might be expected to define, describe and use annotated diagrams or models when answering an examination question (this is similar to Level 1 in the row above). Merit is differentiated by <i>giving reasons</i> for the life process being discussed, and, again as in Level 1, Excellence additionally requires some higher-level thinking such as justifying or evaluating.
91603: Demonstrate understanding of the responses of plants and animals to their external environment (L3)	A slightly more complex description is expected at this level – in addition to showing knowledge of the relevant response process, students must be able to explain the adaptive advantage for that plant or animal in relation to its ecological niche (i.e., two sets of concepts are brought together). Students explain one concept or the other (the response or the adaptive advantage) for Achieved and both concepts for Merit. For Excellence, higher-order thinking (justifying, evaluating, comparing and contrasting) must again be demonstrated in the way the ideas are linked.

“high-achieved” and “low-merit” and “high-merit” and “low-excellence”. It is not clear that examiners are making these judgements based on anything beyond additional recall of declarative knowledge. It is only for Excellence (at all levels) that candidates are required to go beyond simply describing historical phenomena in successively greater detail and demonstrate greater understanding of historical epistemology.

Statistical analyses of predicted progress

Full, national sets of results for achievement standards in history and biology were used to establish achievement scales³ for each discipline. Three scales were established for each discipline, at each NCEA level, for each year included in the analyses: Levels 1 and 2 in 2014 and Levels 2

Table 4. Summary of three externally assessed history achievement standards

Number/ name/ level	Brief commentary on assessment focus/ how achievement is differentiated (A/M/E)
91005: Describe the causes and consequences of an historical event (L1)	Exemplars suggest that Achieved, Merit and Excellence judgements are based on the amount of detail that a candidate could describe. Students gain Achieved if they are able to describe causes and consequences. To gain Merit a description is required to have sufficient detail to demonstrate why an event occurred and the significance of its consequences. Excellence requires candidates to demonstrate a detailed understanding of a topic and establish clear links between its causes and consequences.
91223: Examine causes and consequences of a significant historical event (L2)	At Level 2, candidates <i>examine</i> (rather than just to <i>describe</i>) causes and consequences. With no specified contexts, Achieved, Merit, and Excellence judgements are largely based on the amount of detail that candidates provide to support their answers. In other words, an <i>examination</i> actually seems to mean a more detailed <i>description</i> . Students gain Achieved if they are able to examine causes and consequences and present this in a “meaningful narrative showing historical sequence or chronological order” and Merit if their supporting evidence is detailed. Students gain Excellence if they demonstrate a detailed understanding of the topic as well as some insight in establishing links between causes and consequences.
91438: Analyze the causes and consequences of a significant historical event (L3)	The interpretive, contested nature, and the complexity of the history learning area is apparent at this level and goes beyond mere detailed evidence; it reflects the historical thinking research literature on the procedural concept of <i>cause and consequence</i> . Students gain Achieved if they are able to analyze “underlying and immediate” causes and “short-term and long-term” consequences. Merit requires them to evaluate and prioritize causes and consequences and justify their significance. To gain Excellence students need to demonstrate an understanding of the complexity of causes and consequences.

and 3 in 2015. These three scales reflected performance aggregated across internally assessed standards, across externally assessed standards, and across all standards (both internally and externally assessed).

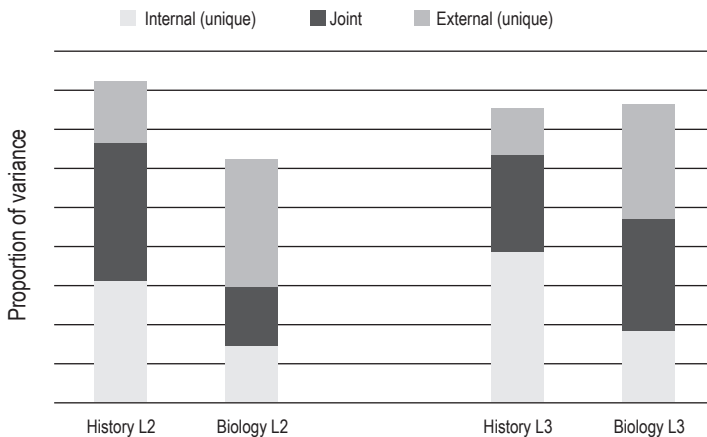
Having established the achievement scales, the extent to which later achievement in a discipline could be predicted from earlier achievement in each of internal and external assessment in the same discipline, was estimated. Figure 1 summarises this information. The *joint* variance component of each bar in the graph represents characteristics of students

that predict later-year achievement and are measured by both internal and external assessment. The *unique* components represent characteristics that also predict later achievement, but which are specifically measured by just internal assessment or just external assessment.

For example, the left-most bar of Figure 1 shows that, altogether, 41% of the variation in achievement in Level 2 history in 2015 (estimated by an achievement scale aggregating performance across internally and externally assessed standards) was accounted for by the previous performance of the same students in Level 1 history, in 2014. Of this 41%, 8% was uniquely associated with variation in external assessment (represented by the top band on the bar), 15% was uniquely associated with variation in internal assessment (bottom band on the bar), and the remaining 18% was jointly associated with variation in internal and external assessment (middle band).

The most striking feature of the analysis is that history and biology are quite different in respect of the assessment modality that most strongly

Figure 1. Proportions of the variance in performance history and biology at Level 2 (leftmost pair of bars) and at Level 3 (rightmost pair of bars) associated with prior-year performance in internal assessment (top panel on each bar), external assessment (bottom panel), and by performance combined across the two assessment formats (middle panel).



predicts later achievement. In history, a substantially larger share of the variance in later achievement is accounted for by earlier achievement in internal assessment than by earlier achievement in external assessment. As noted above, 15% of achievement variance in Level 2 history was accounted for by achievement in internal assessment at Level 1 compared with 8% accounted for by achievement in external assessment at Level 1. The difference was even more pronounced at Level 3: 19% of variance in Level 3 achievement was associated with internal assessment at Level 2, and just 6% with achievement in external assessment at Level 2. In biology this pattern was reversed: 17% of variance in Level 2 achievement was associated with external assessment at Level 1 and 7% with internal assessment at Level 1, with 15% and 6% of Level 3 achievement variance respectively associated with external and internal assessment at Level 2.

Setting aside differences between history and biology in respect of the relative predictive power of internal and external assessment, another observation in relation to Figure 1 is that achievement in Level 1 history was, overall, more predictive of later achievement in Level 2 history than Level 1 biology was of achievement in Level 2 biology. This is evinced by the greater total proportion of variance at Level 2 associated with achievement at Level 1 (41% for history and 31% for biology). This pattern did not hold for Level 2 as a predictor of Level 3 however; for both disciplines, 37% of variance at Level 3 was associated with prior achievement at Level 2.

Discussion and conclusions

In this study we have identified the assessment format associated with the greatest unique component of the variance in later achievement in each of history and biology. Alongside that we have analysed key features of each assessment format in each discipline. Bringing these two parts of the study together, we can begin to form theories of the kinds of learning that will best serve progression in each discipline. For history, the characteristics of students uniquely measured by internal assessment were more strongly predictive of later performance in overall assessment at both Level 2 and Level 3 than aspects uniquely associated with external assessment. For biology, the converse was the case; student characteristics uniquely

measured by external assessment were more strongly predictive of later performance than those uniquely measured by internal assessment.

The qualitative analysis of the requirements of the standards, and assessment practices associated with applying these standards, provides a guide to interpreting the statistical results. The qualitative analysis clearly distinguished the internally and externally assessed history standards; whereas the former were shown to be characterised by a requirement for genuine epistemological thinking, the emphasis in the latter appeared to be more on non-epistemic conceptual knowledge. However, any distinction like this was much less clear in respect of the biology standards. While the internally assessed inquiry standards seemed to show some epistemic focus in that they require the collection of data to support evidence-based investigations, deeper exploration of the achievement criteria, and especially the exemplars, revealed this focus to be superficial. Progress in biological inquiry was characterised by reduction in the level of supervision, better understanding of declarative knowledge, and of theoretical (conceptual) ideas, rather than by increasingly sophisticated epistemic thinking.

These epistemic differences provide a plausible explanation for the greater predictive power of the internally assessed standards in history than of those in biology. Under this interpretation, progress in history is better served by epistemic thinking than is progress in biology. In contrast to history, progress in biology seems to be characterised by the amassing of more declarative knowledge and more sophisticated understanding of existing biological theories, than by the development of the knowledge-building processes by which the knowledge and theories were established. Declarative and conceptual knowledge are arguably most strongly discriminated in formal examination conditions, in which students cannot consult reference materials to aid recall or theoretical (conceptual) exposition. If so, the emphasis on these types of knowledge in the biology standards would explain the greater predictive power of external assessment in biology than in history.

The greater overall power of Level 1 achievement to predict Level 2 achievement in history than in biology suggests that the epistemic focus of the internally assessed standards provides a more coherent structure to learning progress than the aggregation of pieces of externally assessed

knowledge. While the two disciplines show equal power to predict Level 3 achievement from Level 2 achievement, the focus of the biology standards, both internally and externally assessed, suggests that the predictive relationship for biology primarily resides in developing skill in recall of declarative knowledge and in theoretical conceptualisation, rather than in the disciplinary coherence conferred by a stronger grasp of epistemic knowledge.

The findings of this study highlight the way in which the focus of teaching and learning and, potentially, of students' learning motivations, is affected by the content of, and approach to, assessment. The characteristics of students measured by an assessment might relate to the declarative, conceptual, or epistemic knowledge of the discipline itself (e.g., historical thinking), to skills associated with performance in assessments (e.g., writing skill), or to metacognitive factors (e.g., motivation). These different kinds of characteristics combine to influence students' performance in assessments.

In the present study, the later ("predicted") achievement measures are based on assessment results aggregated across the internal and external modalities. Thus assessment-related skills that are primarily associated with one but not both modalities are approximately evenly represented in these aggregate measures. It is therefore unlikely that differences between history and biology in the predictive power of each assessment modality can be accounted for in terms of assessment-related skills. This would leave knowledge-related and metacognitive factors as the most likely mediators of the difference.

In history, the distinctive knowledge-related factor associated with internal assessment, which was a stronger predictor than external assessment of later achievement, was epistemic knowledge. We can only speculate regarding any metacognitive factors that might have contributed to the stronger predictive relationship, but the motivation and engagement provided by taking part in authentic disciplinary inquiry is certainly a plausible possibility. In biology, declarative and conceptual knowledge was the predominant focus of both internal and external assessment. As noted above, a likely reason for the greater predictive power of external assessment is that the strictures of formal examination are likely to result

in stronger discrimination of students' memorisation and application of these kinds of knowledge. Again, any involvement of metacognition in the predictive relationship can only be speculated upon, but motivation to "cram" for examinations seems a likely factor.

In summary, the present research provides a specific illustration of the effect of assessment on the enacted curriculum; when students and teachers understand that the memorisation and application of declarative and conceptual knowledge is what is valued in the assessment this becomes the focus of teaching and learning. On the other hand, when the assessment values the epistemic aspects of a discipline, students are likely to be motivated to demonstrate their capability to develop this kind of knowledge. If the intention expressed in *NZC* to support critical citizenship for the 21st century is to be honoured, its focus on epistemic thinking must be realised; it is this kind of thinking that enables citizens to develop new understandings and to critically evaluate competing explanations of social and physical phenomena. The analyses presented here suggest that this will happen only if New Zealand's high-stakes assessment—NCEA—emphasises this kind of thinking. As the analysis of the history data shows, the internal assessment component of NCEA offers enormous opportunities in this regard.

Notes

1. Some theorists argue that working with models is *the* centrally important epistemic practice in the sciences, and hence should also be central to school inquiry practices (Lehrer & Schauble, 2015).
2. <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/subjects/biology/annotated-exemplars/>
3. The achievement scales are akin to those that underpin widely-used standardised tests such as the Progressive Achievement Tests (New Zealand Council for Educational Research, n.d.) and e-asTTle (Te Kete Ipurangi, n.d.). They are similar to grade point averages (GPAs), but they differ in an important respect: grade point averages are calculated by awarding each student a pre-determined (and arbitrary) number of points for each result (e.g., four points for each Excellence credit, three for each Merit credit, and so on), with the total

score averaged over the total number of credits undertaken to obtain the GPA. On the other hand, the statistical procedure for constructing the achievement scales takes into account the relative difficulty of attaining each result in each standard; unlike GPAs, a given grade is valued differently depending on the difficulty of the standards in which it was attained. Thus, students who do well in more difficult standards attain higher scale locations than those with equivalent grades in easier standards. The relative difficulties of the standards are themselves estimated from the results data by the same statistical model that locates the students on the achievement scale. The statistical model in question, on which the procedure for establishing, or calibrating, the scales is based, is called a one-parameter graded response model (Samejima, 1969).

References

- Barnett, R. (2004). Learning for an unknown future. *Higher Education Research and Development*, 23(3), 247–260.
- Bunnting, C., & Cowie, B., with Anderson, D., Bull, A., Caygill, R., Hipkins, R. ... Vander Zwaag, C. (2017). Towards a systems view of science education in New Zealand. *Curriculum Matters*, 13.
- Colucci-Gray, L., Trowsdale, J., Davies, R., Burnard, P., & Gray, D. (2017). *BERA Research Commission reviewing the potential and challenges of developing STEAM education through creative pedagogies for 21st century learning: How can school curricula be broadened towards a more responsive, dynamic, and inclusive form of education?* doi: 10.13140/RG.2.2.22452.76161
- Hipkins, R. (2012). *Building a science curriculum with an effective nature of science component*. Wellington: New Zealand Council for Educational Research. Retrieved from: <http://www.nzcer.org.nz/research/publications/building-science-curriculum-effective-nature-science-component-0>
- Hipkins, R. (2013). *NCEA one decade on*. Wellington: New Zealand Council for Educational Research. Retrieved from: <http://www.nzcer.org.nz/research/publications/ncea-one-decade>
- Hipkins, R., Bolstad, R., Boyd S., & McDowall, S. (2014). *Key competencies for the future*. Wellington: NZCER Press.
- Hipkins, R., Johnston, M., & Sheehan, M. (2016). *NCEA in context*. Wellington: NZCER Press.
- Lehrer, R., & Schauble, L. (2015). The development of scientific thinking. In R. Lerner (Ed.), *Handbook of child psychology and developmental science* (7th ed., vol. 2). Wiley online.

Building epistemic thinking through disciplinary inquiry

- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington: Learning Media.
- New Zealand Council for Educational Research. (n.d.). *Progressive Achievement Tests*. Available at: <http://www.nzcer.org.nz/tests/pats>
- OECD. (2016). *PISA 2015 assessment and analytical framework. Science, reading, mathematics and financial literacy*. Paris: Author. Available at: <http://www.mecd.gob.es/dctm/incee/internacional/pisa-2015-frameworks.pdf?documentId=0901e72b820fcc48>
- Ormond, B. (2017). Curriculum decisions—the challenges of teacher autonomy over knowledge selection for history, *Journal of Curriculum Studies*, 49(5), 599–619. doi: 10.1080/00220272.2016.1149225
- Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, 25, 177–196.
- Perkins, D. (2009). *Making learning whole: How seven principles of teaching can transform education*. San Francisco: Jossey-Bass.
- Rust, C. (2002). The impact of assessment on student learning. *Active Learning in Higher Education*, 3(2), 145–158.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34(4).
- Seixas, P., & Morton, T. (2012). *The big six: Historical thinking concepts*. Toronto: Nelson Publishers.
- Sheehan, M. (2013a). “History as something to do not just something to learn”: Historical thinking, internal assessment and critical citizenship. *New Zealand Journal of Educational Studies*, 48(2), 69–83.
- Sheehan, M. (2013b). “*Better to do than receive*”: *Learning to think historically through internally assessed course work*. Teaching and Learning Research Initiative (TLRI), January 2011–July 2013. Available at: http://www.tlri.org.nz/sites/default/files/.../Sheehan_Summary_web%20ready.p
- Shemilt, D. (1983). The devil’s locomotive. *History and Theory*, 22(4), 1–18.
- Te Keti Ipurangi. (n.d.). *E-asTTle*. Available at: <https://e-asttle.tki.org.nz/>
- Tytler, R. (2016). Learning progressions from a sociocultural perspective: Response to “Co-constructing cultural landscapes for disciplinary learning in and out of school: The next generation of science standards and learning progressions in action”. *Cultural Studies of Science Education*. doi:10.1007/s11422-016-9777-x
- Zohar, A., & Hipkins, R. (2017, June). How “tight/loose” curriculum dynamics, combined with teachers’ epistemic understanding, impact the treatment of knowledge in two national contexts. Paper presented at Third European Conference for Curriculum Studies, Stirling, Scotland.

The authors

Michael Johnston is a senior lecturer in the School of Education at Victoria University with an interest in the impact of educational assessment on curriculum. He was previously a senior statistician at the New Zealand Qualifications Authority where he worked towards reforms of NCEA between 2005 and 2010. He remains a member of an independent committee of academics providing technical advice to NZQA.

Email: michael.johnston@vuw.ac.nz

Rosemary Hipkins is a chief researcher at NZCER. Rose was actively involved in the development of both *NZC* and NCEA. She has continued to lead national research projects related to curriculum and assessment innovation in New Zealand and has a specific concern with maintaining the fidelity of the curriculum in high-stakes assessments.

Email: rose.hipkins@nzcer.org.nz

Mark Sheehan is a senior lecturer in the School of Education at Victoria University and has been involved in education for over 40 years as a teacher, lecturer, textbook writer, researcher, museum educator, and curriculum designer. His research interests include the place of knowledge and assessment in the secondary curriculum, threshold experiences in teacher education, critical and historical thinking, and the role of history in reconciliation.

Email: mark.sheehan@vuw.ac.nz