

National Standards, moderation challenges and teacher learning

Rosemary Hipkins and Edith Hodgen, NZCER

Paper presented at the Symposium on Assessment and Learner Outcomes, Wellington, September 1-3, 2011

Abstract

Moderation of student work can support teachers to reach a shared understanding of the meaning of a standard, and to more reliably judge a range of evidence in relation to that standard. In theory insights teachers gain via moderation activities could support changes in teaching, leading to improvements in outcomes for students. However moderation has largely been under-researched as a professional learning activity: we need to know more about the dynamics of moderation processes that are successful in supporting professional learning as opposed to those that result in moderation being viewed as simply an accountability-focused demand on teachers' time. NZCER has established a programme of work in this area, starting with a range of questions about school moderation practice in relation to the National Standards in the NZCER National Survey of Primary Schools 2010. This paper will discuss the analysis of teacher survey responses about school moderation practice in relation to the question of its use in professional learning, and consider the implications of the findings in the light of a literature review that has recently been undertaken.

Introduction

In 2010 National Standards were introduced for students in Years 1-8 in New Zealand schools. The standards specify expected progress in reading, writing and mathematics (Ministry of Education, 2009a, 2009b). This initiative was central to the incoming government's education policy following the 2008 elections. Parents were promised there would be regular assessment of primary and intermediate students against the new standards, and that reports to them would be in plain English, so that they could clearly see if their child was making expected progress or falling behind. When the latter was the case, targeted funding would help schools provide more support. The overall aim was to improve levels of achievement in the foundational areas of literacy and

numeracy, and to have all students securely on a learning trajectory that could see them achieve a Level 2 NCEA¹ award in their secondary school years.

Most schools had already been using standardised assessment tools before this initiative was introduced. They can continue to use these as evidence of achievement against the new benchmarks described by the National Standards but they are also expected to combine data from these with a range of evidence from students' work in class. Thus schools and teachers face the challenge of using all the available information to make an overall teacher judgement (OTJ) of individual student performance in terms of the standards. In this way the introduction of the standards requires primary teachers and schools to investigate the *meaning* of the work that students generate. While teachers have always made such judgements informally, moderation as an organised process requires making *collaborative* decisions to reach consensus agreements.

Some very specific conditions need to come together if reporting to parents is to result in positive learning outcomes for children. It is far from self-evident that increased achievement will flow as a benefit from plain language reporting against the standards. Knowledge of how the "New Basics" standards played out in Queensland led two Australian researchers to suggest that standards need to be populated with rich professional insights generated through carefully designed and well resourced moderation procedures. A standard is a *collective* of exemplars, shared experiences and accumulating practice. As such it needs to be owned and tended by the professionals who are expected to be the change agents in education. When teachers do this moderation work together with the explicit intent of supporting students' learning, *pedagogical possibilities* will accumulate in a manner that strengthens practice across the learning collective (Klenowski & Wyatt-Smith, 2010).

With the introduction of National Standards moderation - of pieces of student work and when making OTJs - has become an important professional responsibility for *all* New Zealand's primary school teachers. Many schools already used moderation before the standards initiative began, but it seems likely that many more did not. The 2010 NZCER National Survey of Primary Schools afforded an opportunity to investigate primary principals' and teachers' early responses to this significant policy change, and to investigate relationships between their earlier experiences of moderation and the views they now held of National Standards. Frequency patterns and some cross-tabulations have already been described in the first report from this survey (Wylie & Hodgen, 2010). This paper builds on that report by probing below the surface of the teacher responses to investigate relationships between the responses individuals made to different survey questions.

The survey included several sets of Likert-scaled questions whose items probed teacher views about a range of aspects of moderation, or related issues and views around National Standards. Only teachers who were already moderating student work responded to these particular questions

¹ National Certificate of Education Achievement

– a subset of 769 teachers from the 970 who completed the whole survey. More general questions on national standards were answered by 829 of the 970 respondents (those whose schools had begun at least some work on National Standards). This paper describes seven of the factors revealed by analysis of responses to these banks of items. For reference, these are summarised in Table 1 in the order that they appear in the report.

Table 1 Summary of factors described in this report

| Factor name | Focus of items comprising factor | Cronbach's alpha |
|--------------------------------|--|-------------------------|
| Moderating student work | 4 items: prior experiences of moderating classroom work | $\alpha = 0.78$ |
| Cooperative moderation of OTJs | 4 items: prior experiences of making OTJs collegially | $\alpha = 0.81$ |
| Clarity of standards | 3 items: clarity of NS for making judgements | $\alpha = 0.80$ |
| Sourcing evidence for OTJs | 3 items: ease of sourcing evidence for making OTJs | $\alpha = 0.80$ |
| Assessment changes | 4 items: changes in assessment and reporting practice since NS introduced | $\alpha = 0.67$ |
| Ways of working | 7 items: collegial teacher working and learning related to NS introduction | $\alpha = 0.69$ |
| Positive impact | 5 items: perceptions of direct impacts of NS | $\alpha = 0.84$ |

Our aim in undertaking this further analysis was to better understand how to support ongoing professional learning as moderation activities and decisions become a higher stakes aspect of teachers' work, albeit one that has at least the potential to impact positively on raising student achievement. With that aim in mind, how ready are our primary schools to embrace an idealistic vision of moderation as an ongoing opportunity for professional learning? This report suggests that many teachers arrived at their early encounters with National Standards with a level of resistance arising from a preference for their school's already existing culture of collegial analysis of assessment results, including some level of prior experiences of moderation of student work. Those teachers who had enjoyed limited or no prior experience of working together collegially to monitor student achievement were more likely to report positive changes prompted by the introduction of National Standards.

Evidence of moderation activity before National Standards were introduced

An important decision-making step logically precedes the making of OTJs. First the individual pieces of evidence that will potentially inform the OTJ must be assembled and their meaning *as evidence* determined. To some extent, such meaning is already built into psychometrically scaled

assessment tools such as asTTle and PATs². These types of tools position the achievement of individual students relative to their peers: what the data conveys is a *normative* sense of progress relative to the whole age cohort. The meaning of the overall assessment result has already been made for the teacher by reference to a research-informed scale. Of course there are potentially many layers of additional interpretation under this overall judgement, as teachers determine how and why students responded to the test items as they did.

The National Standards initiative requires schools to use a range of evidence. This is important to note because the meaning of evidence gathered during classroom learning activities brings different professional judgement challenges to interpreting results from nationally standardised tests. Evidence from classroom work can provide more immediate and contextualised insights into the achievements and challenges of individual students so it is important that it be included in any OTJs to be made. Furthermore, teacher responses to the 2010 survey ($n = 769$ for all those teachers who were already moderating work) suggest work gathered during learning is the most common type of evidence in use:

- classroom work (94 percent) and observations (92 percent) were identified as the most common data sources for making OTJs
- assessment tools that require some level of personal teacher judgement (e.g., ARBs, NUMPA)³ were used by 91 percent
- standardised tools such as asTTle and PATs were used by 89 percent
- writing exemplars referenced to MOE benchmarks were used by 71 percent, and referenced to the school's own benchmarks by 62 percent
- Student self assessment (48 percent) and peer assessment (31 percent) were not as commonly accessed as evidence sources, but were being used by some teachers.

Can we reasonably expect all the teachers in a school to make comparable judgements as they convert evidence from individual pieces of student work into data to inform the OTJ process? The literature synthesis we recently completed highlights the potential for differing priorities and personal philosophies to influence the specifics to which teachers direct their attention when determining the meaning of evidence (Hipkins & Robertson, 2011). This question of comparability gains greater urgency when the stakes are higher than simply responding in the moment when interpreting student work to determine next steps for learning. Moderation conversations that develop shared meanings for actual pieces of student work, or for classroom observations, would seem to be particularly important if greater school-wide consistency in

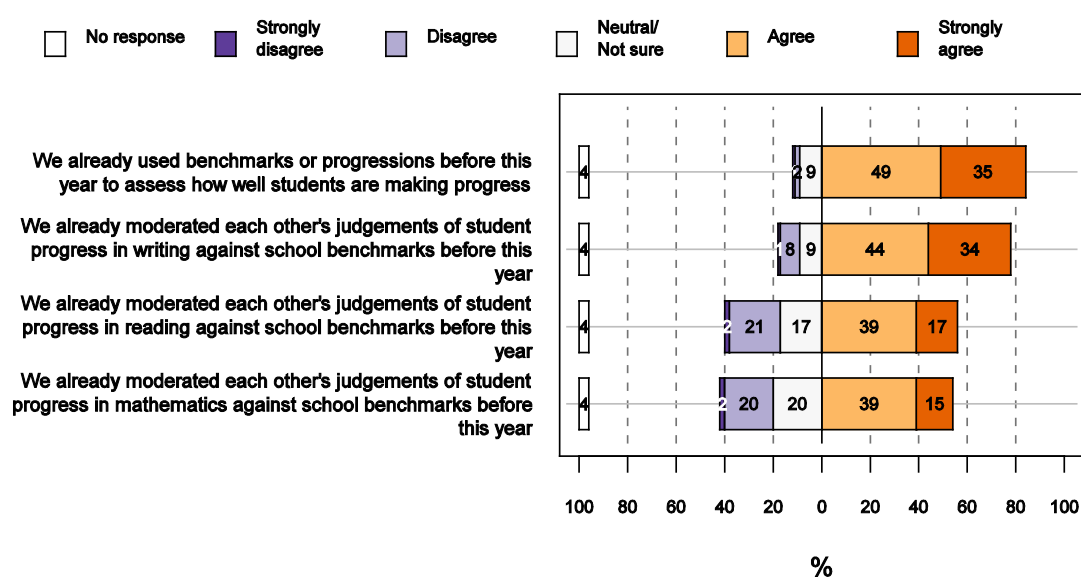
² asTTle = assessment tools for teaching and learning, available for mathematics, reading comprehension and writing; PATs = Progressive Achievement Tests, available for reading comprehension and vocabulary, listening, and mathematics

³ ARB: Assessment Resource Bank items (English, mathematics and science); NUMPA: Numeracy Project Assessment

judgements is to be reached. Addressing this challenge, some schools had already assembled their own benchmarks against which to moderate individual pieces of evidence well before the National Standards initiative began (Hipkins, Cowie, Boyd, Keown, & McGee, 2011).

The factor *Experiences of moderating student work* shown in Figure 1 was drawn from a bank of 9 items that asked about experiences of moderating OTJs. The high Cronbach's Alpha for this factor suggests a relatively high degree of coherence in the way individual teachers answered the items that comprise the factor. As Figure 1 shows, experience of moderation was most likely for writing, followed by reading, and mathematics.

Figure 1 Experiences of moderating student work (n = 769) (α = 0.78)



Average responses for the items that comprise the factor were also calculated. Just 20 percent of teachers had an average score in the range 1.0–1.7, which would be indicative of strongly agreeing with at least three items. This suggests that only a small number of the responding teachers (at the most 20 percent) had prior experience of moderating student work against benchmarks *in all three areas* before the standards were introduced. At the other end of the spectrum of average responses, 19 percent of this sub-group of teachers appeared to have some but very limited experience of moderating student work against benchmarks.

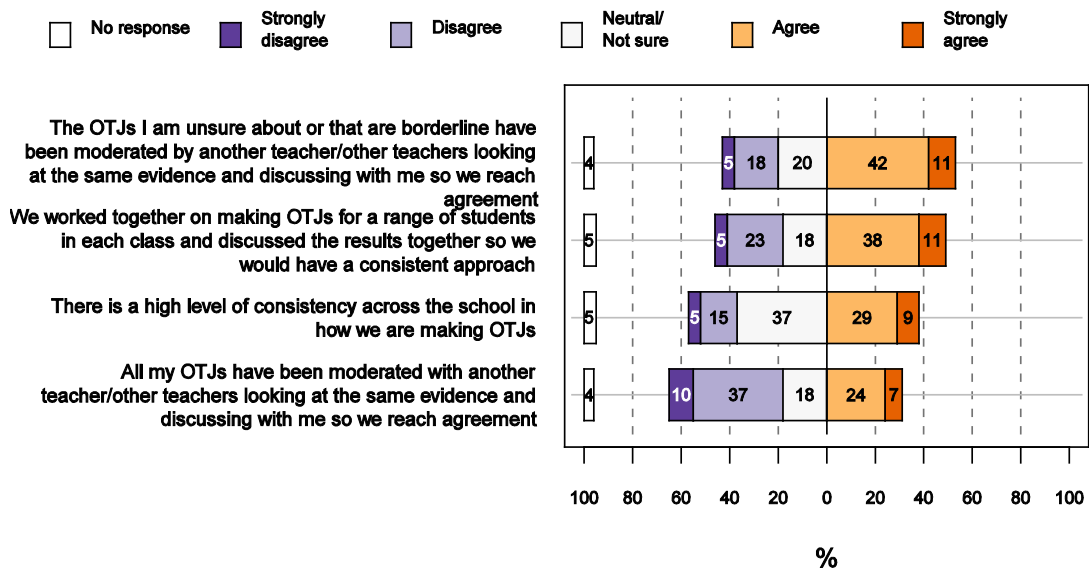
Overall, 43 percent of the whole sample of teachers had not yet had any substantial experience of moderating student work against some sort of benchmark.

Initial experiences in making OTJs

Cooperative experiences of moderating OTJs were probed by the four items that make up the factor called *Cooperative moderation of OTJs* shown in Figure 2. Again the relatively high level

of correlation between the responses from any one individual ($\alpha = .81$) suggests that those who have taken part are likely to have experienced several of these types of experiences. Note that just 38 percent of the teachers thought that there was a high level of consistency across the school when making OTJs and a further 37 percent were unsure.

Figure 2 **Cooperative moderation of OTJs ($n = 769$) ($\alpha = 0.81$)**



Were the teachers who had wider experiences of moderating student work prior to the implementation of National Standards the same group who reported positive experiences of moderation of OTJs? Using Pearson's Correlation Coefficient, we found a moderately strong degree of correlation between the two factors ($r = 0.48$, where $r = 1.00$ would be full correspondence). This suggests that many teachers with strong pre-standards moderation experiences were now taking part in moderation of OTJs, but others were not. The converse also holds: that is, the need to make OTJs introduced some teachers to cooperative moderation practices that they had not previously experienced.

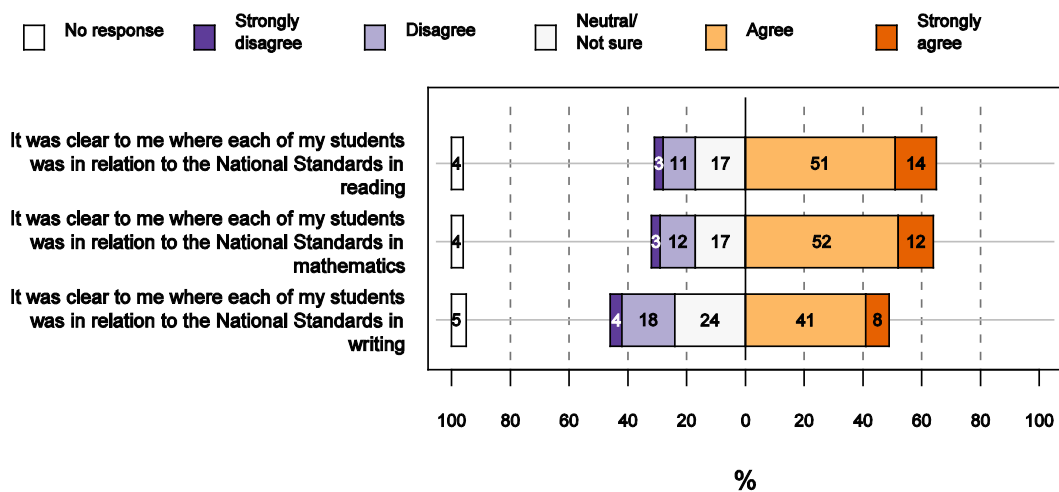
Was the experience of making OTJs productive?

Formal OTJs are initially made to report student progress to parents and to the learners themselves. However if those students identified as not yet meeting a standard are to experience relatively greater gains in their learning (i.e., catch up with their peers) then clearly aspects of the learning they experience will need to change in a way that is supportive of this aim. This logic suggests that any positive impact from National Standards ultimately resides in their usefulness to support schools and teachers to make better decisions about teaching and learning for struggling

students.⁴ As a first step toward this aim, the standards must usefully inform judgements of students' progress. One item set probed this aspect of the standards initiative.

About two thirds of the teachers who had some experience with OTJs reported that it was clear where all their students were in relation to the reading and mathematics standards. Judging writing against the National Standards was not so easy to decide: 49 percent said it was clear where all their students were. These responses formed the factor we called *Clarity of standards* shown in Figure 3.

Figure 3 **Clarity of standards (n = 769) ($\alpha = 0.80$)**

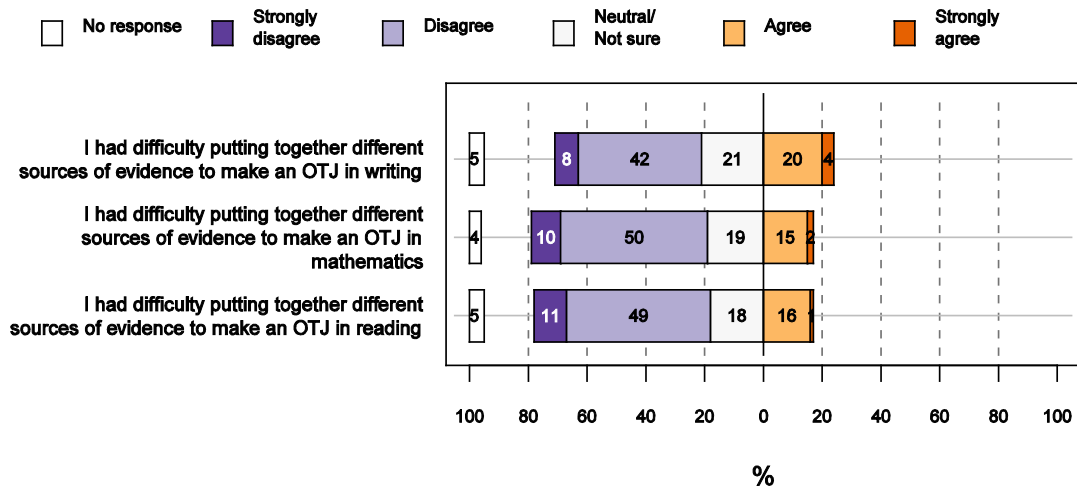


The relatively high degree of coherence of an individual's responses within the factor ($\alpha = .80$) suggests that if teachers found the standards clear in one area they were also likely to find them to be so in two or even all three areas. The skewed spread of average responses showed that relatively more teachers found the standards clear than did not (just 28 percent of average responses were in the 3.0–5.0 range indicating more disagreement than agreement with these items).

Three further items checked to see how easy it had been to gather diverse sources of evidence for the purpose of making an OTJ. These make up the factor that we called *Sourcing Evidence for OTJs* shown in Figure 4.

⁴ Another argument could be that the greater clarity in reporting will galvanise parental concern such that more focused action is taken by the school. But this brings us to the same point: something must then change in the learning opportunities the student experiences.

Figure 4 *Sourcing evidence for OTJs factor* ($n = 769$) ($\alpha = 0.80$)



Average responses for this factor were concentrated at the disagree end of the spectrum but this is a positive response since the items asked about having difficulty. It appears that just 8 percent of teachers or fewer had struggled to assemble evidence in all three target areas.

Are finding the standards clear and not experiencing difficulties in assembling evidence interrelated? Logic suggests that clarity in the standards should help inform likely sources of evidence. However the two factors are only moderately strongly correlated ($r = 0.53$) so there is evidently more to the relationship than this. One possible explanation could be that knowing in theory what type of evidence might help make decisions is not the same as being able to access such evidence in practice. Another possibility is that, having assembled the evidence anticipated to be helpful, the meaning of the standard does not seem clear. This relationship bears further investigation.

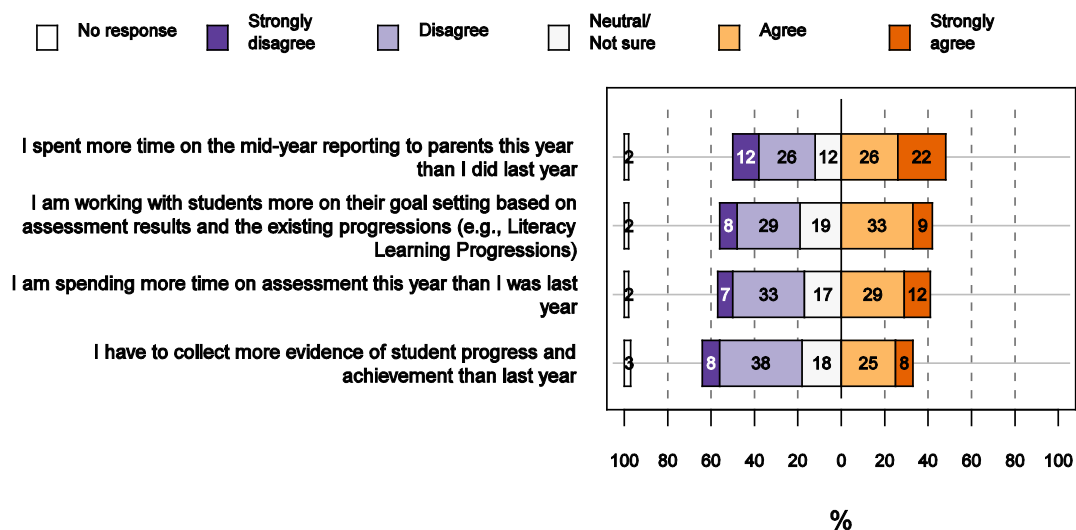
We checked for any pattern of relationships between the two moderation factors introduced above and the *Clarity of standards* and *Source of evidences for OTJs* factors. Cross-tabulation of the different levels of responses for these factors revealed that teachers who strongly disagreed they had difficulty collating sources of evidence (i.e., the positive response) were also more likely to strongly agree that they had already been moderating OTJs cooperatively. Similarly, those who strongly agreed that they could place their students on the standards were more likely to be already moderating OTJs cooperatively. The same pattern held for relationships between the *Experiences of moderating student work* factor and the *Clarity of standards* and *Sources of evidence* factors. Here the relationship is very clear: having already taken part in moderation is associated with finding the meaning of student work clear, and with being able to confidently assemble a range of student work with which to make OTJs.

Changes in ways of working

Moderation could potentially change the ways in which teachers interact with children in response to the insights generated by the additional assessment and reporting-related decisions and activities. From a bank of 20 items that suggested possible changes in teachers' work in response to the National Standards, we identified a factor we called *Assessment changes*. The items that comprise this factor are shown in Figure 5. Note that this factor and the next one to be introduced have somewhat lower reliability levels than the other factors discussed in this report.

While three of the items in the *Assessment Changes* factor could be seen as related to compliance behaviour, it is interesting that one of them is indicative of a clear “pedagogical possibility” (Klenowski & Wyatt-Smith, 2010): 42 percent of the teachers who had taken part in moderation were spending more time working with individual students on setting goals referenced to their specific learning progress.

Figure 5 **Changes in work as a result of National Standards: *Assessment changes* factor ($n = 829$) ($\alpha = 0.67$)**

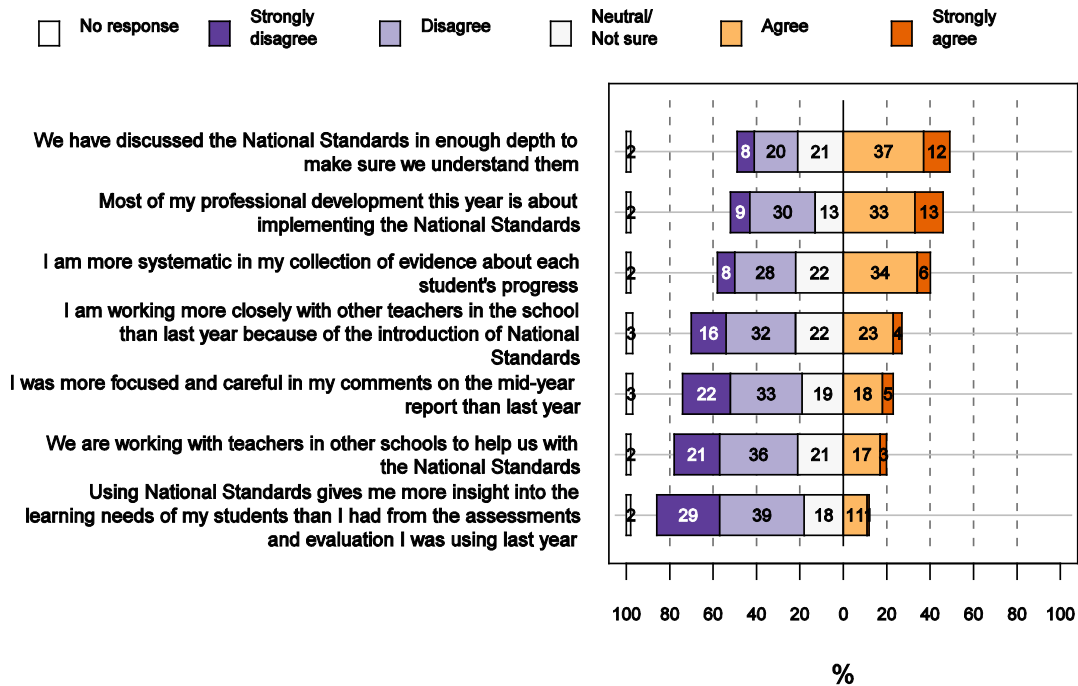


Average responses for this factor were reasonably evenly spread across the continuum of possible responses, with a larger cluster in the middle of the range (2.5–3.0 out of 5).

The item on working with students on goal setting is an interesting inclusion in this factor because it points to a pedagogical possibility with at least the potential to help lift achievement. We found a positive relationship between this goal setting item and the *Moderating student work* factor. Those who strongly agreed they were already moderating student work were more likely to strongly disagree they were now spending more time on goal setting whereas those who had not previously been moderating student work were more likely to agree they were now spending more time on goal setting. The clear implication is that this potentially positive change was associated with not having done this sort of professional work before the standards were introduced.

Another sub-set of seven items from the same bank made up a factor we called *Ways of working*. Whereas the *Assessment changes* items mainly related to changes in the time assessment and reporting activities take, those that comprise this factor are mainly about teacher learning and ways of working together, paying careful focused attention to evidence. Responses for this factor are shown in Figure 6.

Figure 6 **Changes in work as a result of National Standards: *Ways of working* ($n = 829$)**
($\alpha = 0.69$)

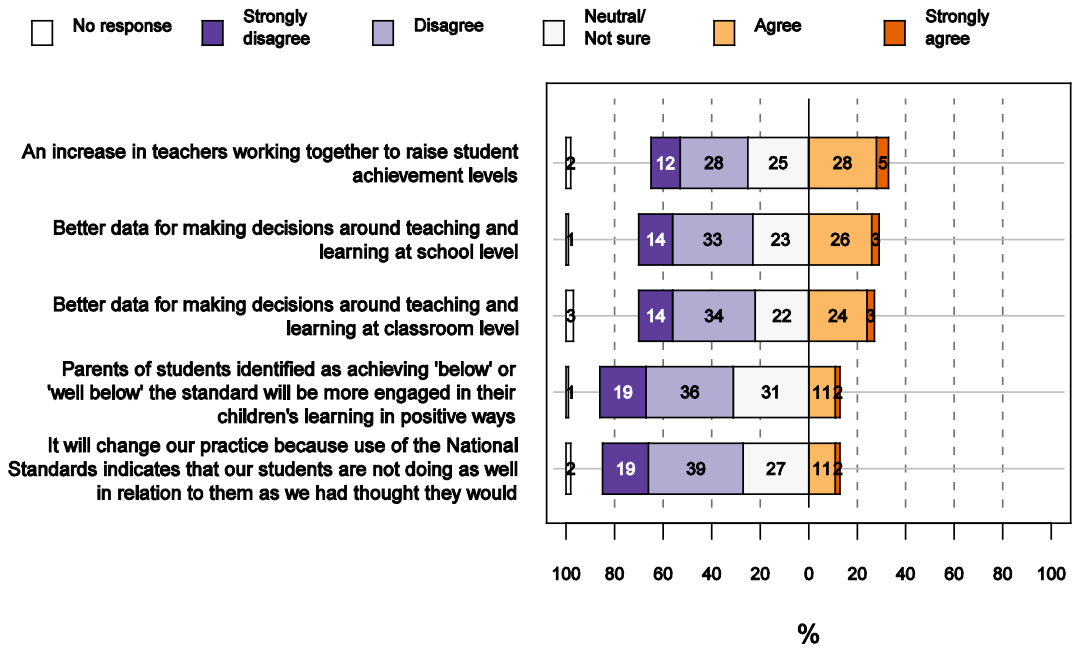


Average responses for this factor were skewed to the negative end of the spectrum. Just 10 percent of teachers agreed or strongly agreed and 31 percent disagreed or strongly disagreed. There was only a moderate degree of correlation between this factor and the *Assessment changes* factor ($r = .41$) where the range of responses was somewhat more evenly spread.

Do teachers see overall benefits for the effort?

One set of items probed teachers' views of the likely short term impact of National Standards in their school. From the 16 possible impacts presented (both positive, negative and equivocal "it depends" type statements), we identified a factor that we called *Positive impact*. This factor is made up of the five items shown in Figure 7.

Figure 7 Short term impacts in school: *Positive impact* factor ($n = 829$) ($\alpha = 0.84$)



Average responses to the items that made up this factor were spread across the spectrum of opinion, but skewed towards disagreement: overall strongly agree or agree, 14%; middle range of responses: 39%; overall disagree or strongly disagree, 45%. There is a clear sense in these negatively skewed responses to the *Positive impact* factor that the effort expended is not perceived by the majority of these teachers to have yielded the intended benefits.

Cross-tabulations revealed a clear pattern of relationships between responses to this factor and responses to the *Assessment changes* factor. Teachers who strongly agreed the impact of National Standards had been positive were also more likely to strongly agree that they had made the changes described in the *Assessment Changes* factor. Recall that this is the group who also appeared to have had little prior experience of moderation of student work. Similarly teachers who strongly agreed with the *Positive impact* factor were more likely to say they were now spending more time on goal setting with students, whereas teachers who strongly disagreed they were spending more time on goal setting were also likely to strongly disagree that the impact had been positive. The picture emerging here is of a small sub-group of teachers (around 50 in the sub-group of 769 who were working with National Standards to moderate student work) for whom the initiative does seem to have opened up new and positive ways of working with evidence of student achievement.

Positive future possibilities?

Any positive impact from National Standards ultimately resides in their ability to help schools and teachers make better decisions about teaching and learning for struggling students by gaining new achievement information to which they did not already have access. This argument appears to assume that teachers (and school leaders) were *not* already endeavouring to access and use achievement data to enhance learning. Other research has found that in schools where the staff had invested considerable time and energy in developing their own benchmarks, and subsequently working together to use these to lift achievement, the National Standards might be seen as an interruption to the school's strong programme of implementation of the New Zealand Curriculum (Hipkins, et al., 2011). Patterns of correlations between the *Positive impact* factor and the other factors introduced in this paper support this earlier research finding. The correlation coefficients are summarised in Table 2, with the factors ordered across the table in the sequence in which they were introduced in this paper.

Table 2 **Correlation coefficients for relationships between the *Positive impact* and other factors**

| | Moderating student work | Cooperative moderation of OTJs | Clarity of standards | Sourcing evidence for OTJs | Assessment changes | Ways of working |
|-----------------|-------------------------|--------------------------------|----------------------|----------------------------|--------------------|-----------------|
| Positive Impact | -0.15 | -0.01 | 0.19 | 0.04 | 0.29 | 0.57 |

Notice the low and negative association between perceptions of a positive impact and earlier experiences of moderating student work and making OTJs. The association between ease of using the standards to make moderation decisions and the Positive impact factor is also low. Cross-tabulations of the *Positive impact* and *Clarity of Standards* factors reveals that teachers with higher levels of confidence in their abilities to assemble evidence and make judgements are also the least likely to perceive positive benefits in the introduction of National Standards. However there is a large group of teachers sitting in the middle ground for this particular cross-tabulation – seemingly indicating some level of uncertainty, both about their abilities and about the potential benefits of using the National Standards. This group must constitute a space of possibility for those charged with working with teachers to effect positive change.

The stronger association between the *Positive impacts* and *Ways of working* factors is a two-edged sword. Cross-tabulations reveal that for those who have made changes in ways of working with other teachers, the impacts and benefits have flowed on. However, as we have seen, teachers who were already working in these ways were more likely to disagree that the impact of National Standards has been positive.

Overall then, where strong change in assessment practice was needed it does seem to have happened – at least in some schools. This is the “good news” in the data. However where

assessment and moderation practice was already strong the National Standards seem to have succeeded only in fostering a degree of negativity in teachers. This is a challenge because perceptions of strong pre-existing practice may or may not be justified. Where further strengthening of practice does prove to be needed (but not yet perceived in the school) it might have become that much harder to achieve, with resistance to be overcome before the staff can move on. The real space of possibility would appear to relate to the many teachers “in the middle” who appeared to not yet have formed clear views of the National Standards, or made National Standards related changes in practice.

References

- Hipkins, R., Cowie, B., Boyd, S., Keown, P., & McGee, C. (2011). Curriculum Implementation Exploratory Studies 2. Final Report: February 2011. Wellington: Ministry of Education.
- Hipkins, R., & Robertson, S. (2011). Moderation and Teacher Learning: What can research tell us about their interrelationships? Wellington: New Zealand Council for Educational Research.
- Klenowski, V., & Wyatt-Smith, C. (2010). Standards, teacher judgement and moderation in contexts of national curriculum and assessment reform. *Assessment Matters*, 2, 107-131.
- Ministry of Education. (2009a). Mathematics Standards for Years 1-8. Wellington: Learning Media.
- Ministry of Education. (2009b). Reading and Writing Standards for Years 1-8. Wellington: Learning Media.
- Wylie, C., & Hodgen, E. (2010). NZCER 2010 Primary & Intermediate Schools National Survey: A snapshot of overall patterns and findings related to National Standards. Wellington: New Zealand Council for Educational Research.