

DO SUBJECT SPECIALISTS PRODUCE MORE USEFUL FEEDBACK THAN NON-SPECIALISTS WHEN OBSERVING MATHEMATICS LESSONS?

Sheila Evans¹, Ian Jones², Clare Dawson¹

¹University of Nottingham, ²Loughborough University

Schools, districts and inspectorates routinely use non-specialists to observe lessons for accountability and professional development purposes. However, there is little empirical research on how well non-specialists observe lessons. We describe two pilot studies in which education professionals made judgements about mathematics lesson observation reports, written by both specialists and non-specialists. In terms of providing feedback to the observed teachers, the professionals considered the specialists' reports to be significantly more useful than the non-specialists' reports. Written advice about a teacher's practice influenced these judgements. The paper considers theoretical and practical implications, as well as limitations of our findings.

Lesson observations are common practice around the world for the evaluation and professional development of school teachers (Lewis, Perry, & Murata, 2006; Ofsted, 2012). They provide an opportunity to improve practice and can influence a teacher's career or a school's status. Many of these observations are conducted by teachers who are not specialists in the subject being taught (Wragg, Wilkley, Wragg & Haynes, 2002). The research reported in this article was prompted by an intuitive assumption that subject specialists are better positioned than non-specialists to give feedback on observed lessons, along with a paucity of research as to whether this assumption is warranted.

One notable study that did touch upon the role of subject specialism when observing lessons was conducted by Wragg *et al.* (2002). Using questionnaires and case studies, the researchers found that teachers often judge observation feedback most helpful to improving practice when the lesson observation was conducted by a subject specialist. Where the observer was not a subject specialist feedback was "bereft of ideas [on how to improve the lesson]" (p. 200) and could be "bland [when the observer] did not have first-hand experience of the subject" (p. 203).

A later study by Peake (2006) provided further support to the importance of subject expertise. Peake, using questionnaire- and survey-based methods, found that teachers working in post-compulsory education considered subject-specialist observers to offer substantially more helpful feedback than non-specialists. Moreover, some teachers were inclined not to take feedback seriously from non-specialist observers.

We have encountered no studies beyond Wragg and Peake in which the subject specialism of the observer is a concern. Instead the research focus is typically on student learning gains (Strong, Gargani & Hacifazlioglu, 2011) and the development

of lesson observation protocols, methods and skills for research purposes (Douglas, 2009). Nevertheless, a theme within this literature is that professional knowledge and experience appears to impact on what is noticed and prioritised when observing lessons (Grant, Hiebert & Wearne, 1998). Furthermore, the literature is clear that what teachers perceive as useful in an observation report depends on their expertise (Carter, Cushing, Sabers, Stein & Berliner, 1988; Colestock & Sherin, 2009; Santagata, Zannoni & Stigler, 2007; Star & Strickland, 2008). For instance, a novice teacher may find advice on classroom management more useful than the subtleties of dealing with unanticipated misconceptions. Conversely, it is these very subtleties that concern expert teachers.

To our knowledge there are no studies that directly test the qualitative hypotheses drawn-up by Wragg and Peak. We conducted two studies to help address this gap. We first investigated whether subject specialists produce written lesson observation reports that (i) are distinguishable from those of non-specialists, and (ii) are more “useful” in terms of helping a teacher improve her teaching compared to those of non-specialists. Integral to this study is the exploration of participants’ understanding of “useful feedback”.

OBSERVED LESSONS

Two experienced mathematics teachers taught four lessons in a UK secondary school. One teacher taught two lessons with a class of 12 and 13 year olds and the other with a class of 15 and 16 year olds. Two teachers, one specialist (mathematics) and one non-specialist (English language) observed each lesson. In total, four observers observed two lessons each. Each observer completed an unstructured report framed by questions based on typical observation forms: *What is your overall impression of the lesson? What is the lesson about? How did student learning take place? How could the lesson be improved?* The completed reports were anonymised and the subject specialism of the observer was not indicated on the reports.

In common with the majority of routine observations, all observers were known to the teachers; they were colleagues. It was assumed that the specialists knew more about, and shared more of each teacher's beliefs, style of teaching, issues and goals.

In a traditional lesson, students often work on an exercise using the same method. Student misconceptions, difficulties and errors are predictable. In contrast, the lessons in this study were based around non-routine, unstructured tasks. These lessons can proceed in unexpected ways; students can use unanticipated solution-methods and unforeseen difficulties including misconceptions may arise. We predicted that compared to a more traditional lesson, these lessons would provide greater opportunities for observers to suggest feedback to help improve teacher practice. For instance, advice on how to help students make connections between various solution-methods. This in turn, may draw out the differences between reports written by the specialist and non-specialist observers.

In accordance with the literature, we expected all observers would provide general pedagogical advice, but only subject specialists would provide advice that draws on their pedagogic content knowledge and their subject knowledge (Shulman, 1986). For instance, all observers may provide advice on student engagement, but only the specialist observer would provide advice on how to orchestrate a whole class discussion in order to build on the collective sense-making of the students.

STUDY 1

The purpose of Study 1 was to establish whether the lesson observation reports produced by specialists were distinguishable from those produced by non-specialists. Twelve professionals, namely teachers (6), teacher educators (4) and researchers with teaching experience (2) drawn from a range of specialisms (art, general education, geography, German, history, mathematics) participated in Study 1.

The observation reports were divided into four sets of four reports such that no set contained more than one report written by a given observer. Each participant received one set of reports. The task of the participants was to decide whether a mathematics or English language specialist had written each report. Participants could also write a comment about each decision. In total each report was independently categorised six times.

Nine of the twelve participants correctly categorised all four of their allocated reports as having been written by specialists or non-specialists. A further two participants correctly categorised just two reports. The remaining participant incorrectly categorised all four reports.

To test whether the twelve participants as a whole categorised the eight reports at a level above chance we conducted a Mann-Whitney U test, comparing our group of participants with a hypothetical group of twelve participants performing at chance. The result demonstrated that the participants were indeed able to correctly categorise the reports at above chance level ($z = -3.20, p < .01$).

The comments provided by the participants revealed that the most common basis for deciding whether to categorise a report as produced by a specialist or not was the degree and sophistication of mathematical content. For example, one participant correctly categorised a specialist observation and wrote, "The type of observer is given away at the end by the statement ' $\sin x = 0.5$ has infinite solutions but is not always true'. Would an English language specialist be able to comment like this?" Conversely, another participant correctly categorised a non-specialist report because of its lack of mathematical content.

STUDY 2

The purpose of this second study was to establish whether specialists' observation reports were perceived as more useful in terms of helping the observed teachers improve their practice, than those of the non-specialists. Subsequently, their

understanding of “useful feedback” was explored. It was likely that teachers would know the authors of the reports. This knowledge could influence their judgments. For instance if they knew the Head of Mathematics wrote a report then they may assume the report was useful. Evaluation therefore might depend more on who has written the report rather than whether or not it was a worthy one. So, instead of asking the teachers to judge the reports, eight mathematics education professionals, namely teacher educators (2), researchers with teaching experience (6) participated. None had participated in Study 1. These participants did not know the teachers; they did not know whether they were novices or experts. Their judgments were based on the reports alone; not whether the advice matched the expertise of the teacher.

A comparative judgement method (Thurstone, 1927) was used to rank the lesson observation reports in terms of perceived usefulness as feedback to the observed teachers. The outcome of the pairwise judgements can then be used to construct a psychological scale of artefacts from “best” to “worst” (Bramley, 2007).

Each participant was presented with eight pairs of reports and asked to decide, for each pair, which report they thought provided the most useful feedback to the observed teacher. In total, every possible pairing of observation reports was judged twice, each time by a different participant, resulting in 56 pairwise judgements. Once the judgments were complete, participants were asked to comment on their decisions.

We independently coded each report; categorising “suggestions for improvement” as being based on either (i) general pedagogic knowledge, (ii) pedagogic subject knowledge or (iii) subject knowledge. To gain further insight into the types of advice prioritised by observers we drew on Wake’s (2011) work on knowledge for teaching and learning. We subdivided the pedagogic subject knowledge and subject knowledge into six categories of subject knowledge for teaching (Ball, Thames and Phelps 2008). This may clarify what is valued in an observation report.

ANALYSIS AND RESULTS

The participants’ pairwise judgments were statistically modelled (Bramley, 2007) to produce a parameter estimate and standard error for each report. These parameters enabled the construction of a scaled rank order of reports from “best” to “worst”, as shown in Figure 1. The top four reports were those by the specialist observers (labelled “S”). The internal consistency (Rasch Separation Reliability (Bramley, 2007)) for the scaled rank order was .65, an acceptably high reliability for discriminating between two groups (specialist and non-specialist).

To investigate these groupings further, we categorised each lesson observation report as either in the top half (assigned a value of 1) or the bottom half (assigned a value of 0) of the rank order. Fisher’s exact test using “specialism” and “top or bottom” as categorical variables reached significance ($p = .029$, two tailed), supporting interpreting the result as two distinct groups of four reports. Study 2 therefore provided support that the participants perceived the specialists’ reports to be more useful in terms of feedback to the observed teachers than the non-specialists’ reports.

Participant feedback

All eight participants cited a preference for reports that made concrete suggestions for improvement. However, beyond this there was no clear consensus as to what constituted a more “useful” report. For example, some cited a preference for reports that described the lesson in detail whereas others had a preference for reports that avoided detailed description. Surprisingly, only two participants explicitly cited mathematical content as influencing their judgement decisions.

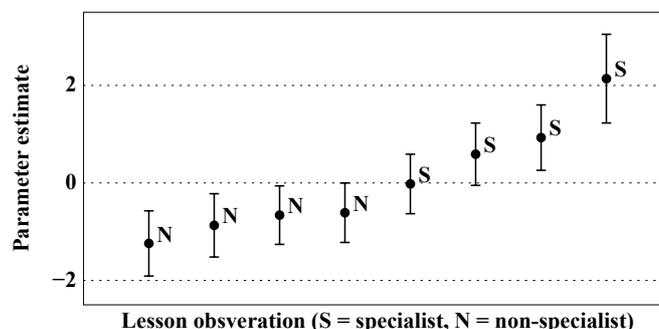


Figure 1: Scaled rank order of the lesson observation reports.

Coding observation reports

Overall, there was consistency between the authors’ coding. The specialists offered a total of 22 suggestions for improvement, ten of which drew on subject knowledge, the non-specialists offered a total of five suggestions, all drawing on general pedagogical knowledge. Table 1 shows the ten math-based suggestions categorized, using a summarised version of Ball, Thames and Phelps’ (2008) categories of subject knowledge for teaching.

<i>Category</i>	<i>Description</i>	<i>Suggestions</i>
Specialised Knowledge SCK	Content Mathematical knowledge unique to teaching	2 explicit 3 implicit
Common Knowledge CCK	Content Mathematical knowledge and skills, not unique to teaching	1 explicit 2 implicit
Horizon Knowledge HCK	Content Understanding how to develop and build on students current knowledge	0 explicit 0 implicit
Content of Knowledge and Students CKS	Understanding how groups of students talk about and handle specific tasks	4 explicit 1 implicit
Content of Knowledge and Teaching CKT	Understanding the design of teaching tasks/sequences of instruction	2 explicit 1 implicit
Content of Knowledge and Curriculum CKC	Understanding how the lesson relates to the curriculum and assessments	1 explicit 0 implicit

Table 1: Categorized numbers of “suggestions for improvement” in the reports.

The authors noted some reports contained additional observer comments, that although not explicitly advice, could be construed as potentially helpful to teachers, especially if they intended to re-use the lesson. For example, an SCK comment: “pairs did not get to grips with Tanya’s method. No one spotted that her lines were drawn wrongly, or that she was wrong to assume that one particular vertex was optimal”.

Although observers did not teach the students Mathematics, there were five instances of the use of the CKS domain. On these occasions subject specialists noticed, in the moment of observing, and subsequently reported on, how students were talking about the mathematics and handling the challenges of the task. For example, one observer stated “The questions: What assumptions did they make? Were they valid? Was their mathematics correct? seemed a bit hard even for this bright group”. Only one observer suggestion was based on how the lesson relates to the CKC domain. Considering the lessons were non-standard and the pressure for students to achieve in high-stake, content driven tests, this is surprising.

GENERAL DISCUSSION

The participants in Study 1 correctly distinguished lesson observation reports written by specialist teachers from those written by non-specialist teachers. The presence or absence of mathematical content appeared to be the key discriminator between the reports. The participants in Study 2 perceived that lesson observation reports written by specialists were more useful in terms of helping teachers improve their practice than those written by non-specialists. These judgements were not based on the presence or absence of mathematical content, but the presence of suggestions for improvement. The authors’ coding of the reports corroborated this. Specialists offered substantially more suggestions than the non-specialists. However, although participants tended not to explicitly refer to the mathematical content of these suggestions, nearly half the specialist suggestions drew on subject knowledge, whereas non-specialists provided no mathematics-based advice. Surprisingly, nearly half these mathematics-based suggestions were based on the CKS domain. We conjecture that the teachers are drawing on their own knowledge of students when noticing and evaluating how students are progressing with a task.

Limitations

The materials were drawn from just four observers, four non-standard lessons and two mathematics teachers, all from one school. Caution must therefore be exercised as to the generalisability to other teachers, lessons, schools and subject areas. The finding from Study 2 generalises only to the study participants. That is, we expect that the same group of participants would perceive specialist reports to be more “useful” than non-specialist reports in general. However, we cannot generalise beyond this group of participants to expect that all mathematics education professionals would perceive observation reports similarly. Results may be quite different if, for example the observed teachers were all novices or the lessons were of a more traditional structure and content.

Theoretical implications

What is it about a specialist teacher’s lesson observation report that mathematics professionals perceive to be more useful than a report of a non-specialist? Study 1 suggests that a key discernible difference is the presence of mathematical content.

However mathematical content was not cited at all by six of the eight mathematics professionals in Study 2, who nevertheless preferred the specialists' reports. One possible explanation is simply that subject specialists are better at providing useful feedback. Participants may respond more positively to reports by members of the same community, mathematics education, as they are likely to share similar beliefs, values and goals. Furthermore, the study showed that their reports did indeed provide more pedagogical advice whether of a general or specialist nature. If this is the case then we should expect the result of Study 2 to generalise to other subject disciplines. For example, we would expect history teachers to produce history lesson observation reports perceived as more useful than those produced by teachers of other subjects.

Another possible explanation is that mathematics teachers are simply better at producing useful feedback than language teachers per se, rather than just for the case of lessons in their own discipline. Although this is a provocative hypothesis, studying mathematics is widely regarded to increase general analytic skills (e.g. Smith, 2004), which might include lesson observation skills. If mathematics teachers are indeed generally better at observing any lesson than non-mathematicians then the finding from Study 2 would not be expected to generalise to other subject specialisms. For example, if the study were reversed so that mathematics and language teachers observed language lessons, then we would not expect the subject specialists' reports (language teachers in this case) to be perceived as more useful than the non-specialists' reports.

Conversely, the paucity of advice offered by non-specialists may be explained by the widely held belief that mathematics is a 'difficult' subject. Non-specialists may lack the confidence to offer advice to mathematics teachers. If this is the case, then only when observing mathematics lessons, and perhaps other technically demanding subjects, would it be perceived that specialists offer more useful advice than non-subject specialists.

We are currently undertaking further research to address the above limitations, and to discern between these possible explanations.

Practical implications

If it is the case that mathematics teachers are "better" at observing mathematics lessons than non-specialists, or that non-specialists do not feel equipped to offer advice, then the practical implications are self-evident. Lesson observations are commonly used for professional development and accountability purposes, and it is vital that they are of high quality. However it is standard practice in many countries for high-stakes observations of mathematics lessons to be conducted by non-specialists. The findings reported here contribute some evidence that schools, districts and inspectorates might be advised to ensure that lesson observations, when intended to help mathematics teachers develop their practice, involve mathematics subject specialists whenever possible.

References

- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407.
- Bramley, T. (2007). Paired comparison methods. In P. Newton, J.-A. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards* (pp. 264-294). London: QCA.
- Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Professional-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education*, 39, 25-31.
- Colestock, A., & Sherin, M. (2009). Teachers' sense-making strategies while watching video of mathematics instruction. *Journal of Technology and Teacher Education*, 17, 7-29.
- Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. *Educational Researcher*, 38, 518-521.
- Grant, T. J., Hiebert, J., & Wearne, D. (1998). Observing and teaching reform-minded lessons: What do teachers see? *Journal of Mathematics Teacher Education*, 1, 217-236.
- Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher*, 35, 3-14.
- Ofsted. (2012). *School inspection handbook: Handbook for inspecting schools in England under section 5 of the Education Act 2005*. London: Office for Standards in Education.
- Peake, G. (2006). *Observation of the practice of teaching*. Huddersfield, UK: PCET Consortium, University of Huddersfield.
- Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, 10, 123-140.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Smith, A. (2004). *Making mathematics count*. London: The Stationary Office.
- Star, J. R., & Strickland, S. K. (2008). Learning to observe: using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education*, 11, 107-125.
- Strong, M., Gargani, J., & Hacifazlioglu, Ö. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, 62, 367-382.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273-286.
- Wake, G.D. (2011). Teachers' professional learning: Modelling at the boundaries. In G. Kaiser, W. Blum, R. Borromeo Ferri, & G. Stillman (Eds.), *Trends in teaching and learning of mathematical modelling* (pp. 653-662). The Netherlands: Springer.
- Wragg, E. C., Wilkley, F., Wragg, C. M., & Haynes, G. S. (2002). *Teacher appraisal observed*. London: Routledge.