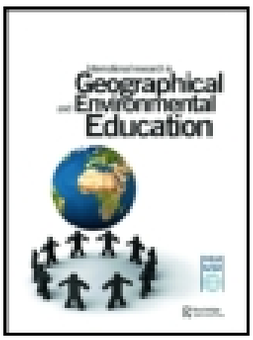

International Research in Geographical and Environmental Education

ISSN: 1038-2046 (Print) 1747-7611 (Online) Journal homepage: http://www.tandfonline.com/loi/rgee20

Assessment in geography education: a systematic review

Rod Lane & Terri Bourke

To cite this article: Rod Lane & Terri Bourke (2017): Assessment in geography education: a systematic review, International Research in Geographical and Environmental Education, DOI: 10.1080/10382046.2017.1385348

To link to this article: http://dx.doi.org/10.1080/10382046.2017.1385348

INTERNATIONAL RESEARCH IN GEOGRAPHICAL AND ENVIRONMENTAL EDUCATION, 2017
https://doi.org/10.1080/10382046.2017.1385348

Assessment in geography education: a systematic review

Rod Lane [a] and Terri Bourke [b]

[a]Department of Educational Studies, Faculty of Human Sciences, Macquarie University, Sydney, Australia; [b]Faculty of Education, Queensland University of Technology, Brisbane, Australia

ABSTRACT

There are more than 700 articles exploring assessment in geography education. However, these papers vary in the degree to which recommendations and conclusions are based on research evidence. Globally, evidence-based practice is being prioritised, making it essential to understand exactly what the empirical research around this topic is saying. A systematic review provides a rigorous method for achieving such a task. This paper quantifies and systematically accounts for the proportion and scope of articles dedicated to assessment in geography education. We conclude that clarity is required regarding: (1) the essential geographical knowledge and skills students should develop; (2) the nature of the learning progressions in each of these areas; and (3) the types and formats of assessment instruments that will provide valid and reliable measures of this progress.

KEYWORDS

Assessment; geography education; geographical knowledge; systematic review

Introduction

There is a general consensus amongst researchers that both formative and summative assessment is integral to improving geography education (Lambert, 2011; Weeden & Lambert, 2006). More broadly, Newton (2007) maintains that the key aim of assessment is to drive improvements in student learning by providing valid, reliable and timely data to inform key stakeholders including teachers and policy makers. To achieve this, clarity is needed about the best methods for capturing student learning information in this domain. While we acknowledge that studies have addressed some aspects of assessment in geography education, a systematic review of the empirical research has not been conducted to provide an evidence base for effective practice. In this paper, we address this need by quantifying and systematically accounting for the proportion and scope of empirical articles dedicated to the assessment of knowledge and skills performance in geography.

There are more than 700 articles exploring assessment in geography education which can be divided into discussion/non-empirical papers and papers based on primary data collected in a geographical context. The themes addressed in the non-empirical papers include, for example, assessment for learning (Lambert, 2011; Weeden & Lambert, 2006); controlled assessment (Atherton, 2011); learning progressions in geography (Bennetts,

CONTACT Rod Lane rod.lane@mq.edu.au
Supplemental data for this article can be accessed at https://doi.org/10.1080/10382046.2017.1385348.

2005a, 2005b; Edelson, Shavelson, & Wertheim, 2013; Solem, Huynh, & Boehm, 2015); and rubric design and success criteria (George, Clarke, Davies, & Durbin, 2002; Marcello, 2009; Smothers, 2002). Other papers have reviewed general assessment issues in geography education (see Bettis, 2001; Weeden, 2013). There have also been major reports published, the most recent of which includes *A road map for 21st century geography education: Assessment* (Edelson et al., 2013)[1] which explores how changes in assessment practices can support efforts to improve K–12 geography education. These publications and reports make an important contribution to our understanding of the general principles of assessment. However, with the global focus on evidence-based practice it is essential to understand exactly what the empirical research around this topic is saying. This paper therefore systematically reviews the empirical research on assessment in geography education, synthesising current knowledge on evidence-based practice in order to highlight gaps and areas for future research.

After explicating the detailed method for this study including the search strategy, criteria and synthesis approach, we provide an outline of the general characteristics of the research including publisher, location, scope, program level and research design. The review then examines the key themes in the literature in order to address two key questions: (1) what do we know about evidence-based assessment in geographical education? and (2) what gaps are identified to point to future research directions?

Method

Systematic reviews are common in healthcare professions such as psychology, nursing, public health, occupational therapy, speech therapy, and physiotherapy but are also increasing in other academic fields such as sociology, business management and education. Systematic reviews provide robust and reliable summaries of a given topic which can be used to inform policy (Petticrew & Roberts, 2006) or from which other conclusions can be drawn and decisions made (Antman, Lau, Kupelnick, Mosteller, & Chalmers, 1992; Oxman & Guyatt, 1993). As well as their usefulness in documenting what existing literature says, they also identify gaps pointing to future research directions (Petticrew & Roberts, 2006). In the last decade, several systematic reviews in the discipline of geography have been undertaken; for example, reviews on global environmental and climate change (Ford & Pearce, 2010; Rudel, 2008), risk assessments and vulnerability (Plummer, de Loë, & Armitage, 2012) and droughts (Tánago Gonzales, Urquijo, Blauhut, Villarroya, & De Stefano, 2016). However, in geography education there appears to be less focus on employing this technique for reviewing literature.

Systematic reviews use explicit eligibility criteria and are conducted according to a predefined methodological approach or protocol which can be replicated (Moher et al., 2015). In other words, the search approach is designed by establishing inclusion and exclusion criteria to determine the selection of studies to be reviewed. These studies are then themed or classified/coded (Hofmann, Hinkel, & Wrobel, 2011) before they are analysed following specific criteria which can be statistical, descriptive or qualitative. The protocol ensures that a systematic review is carefully planned in advance, thus promoting a consistent approach to the review of the literature. Here, we outline the protocol in terms of the search strategy, criteria and synthesis approach.

Search strategy, criteria and approach

The first step was to identify appropriate search terms and parameters. We used six databases relevant to curriculum and assessment to search for articles that met the following criteria:

- Empirical articles defined as meeting the following criteria: (1) a clear statement of the aims of the research/issue being investigated; (2) study placed in the context of existing knowledge, theory, policy or professional practice; (3) clear description of participants and methods; (4) primary evidence, such as interview responses or survey data, in addition to description and analysis of the research findings; and (5) response rate stated, in the case of sample surveys.
- Peer-reviewed papers in English published between 2000 and 2016. We chose to investigate only those articles published during the 16-year period to ensure currency of research and to enable emerging trends in assessment in geography education to manifest. We restricted the search to peer-reviewed journal articles because of the challenges in identifying and obtaining a complete set of other types of publications such as conference papers, monographs, books and institutional reports (Murray, Nuttall, & Mitchell, 2008).
- Abstracts or keywords referred to "secondary school*", or "high school*", or "middle school*", or "primary school*", or "K–6", or "K–12", or "secondary education" or "primary education"
- Abstracts and keywords included "assessment" or "evaluation", and "geography"
- We excluded articles investigating post-secondary, higher or tertiary education and papers not explicitly focused on the assessment of geography knowledge and skills.

Having conducted this search, an additional survey of the main national and international journals in environmental and geography education was undertaken to ensure pertinent articles were not missed. We targeted the following journals: *Environmental Educational Research*; *Geography*; *Geographical Education*; *International Research in Geographical and Environmental Education*; *Journal of Geography*; *Journal of Geoscience Education*; and *Teaching Geography*. To further ascertain complete coverage, we cross referenced the bibliographies of published books in the area of geography education/assessment.

Having retrieved the articles (n = 701), we individually reviewed papers to check for alignment with the above criteria. The outcome of this review process is shown in the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) (Moher, Liberati, Tetzlaff, & Altman, 2009) diagram (Figure 1) which provides a graphical representation of papers identified and reviewed from each database.

Synthesis approach

After identification of the items for analysis, we scanned the titles and abstracts (superficial examination) and set up a log that documented the articles' bibliographical details, search database, program level (primary or secondary), scope of the investigation (local, regional, global/single site, multi-site, international), location and research methods

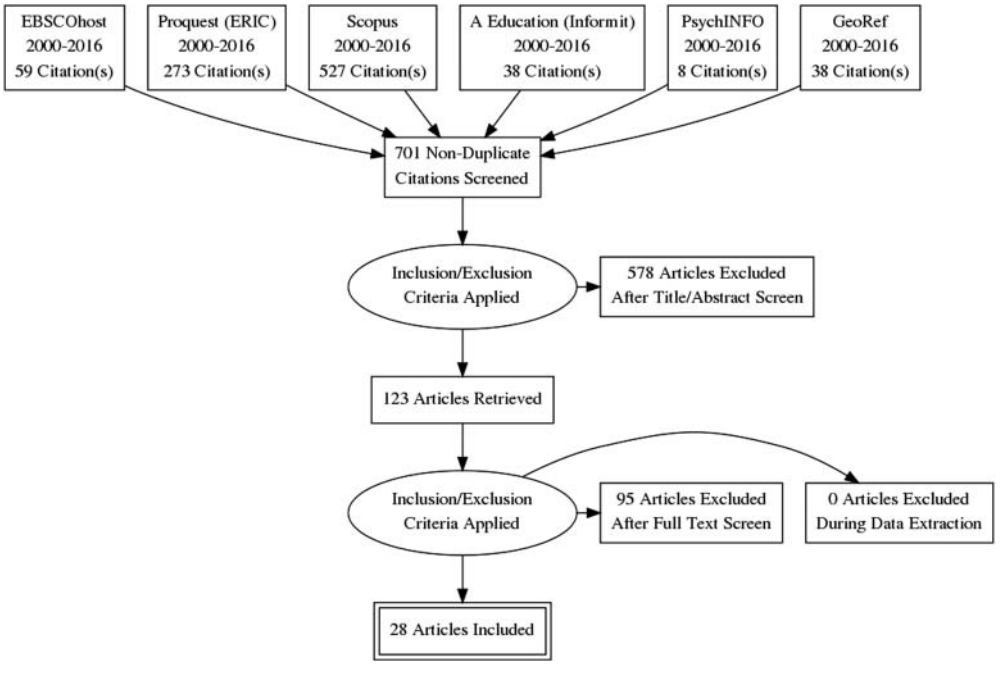

Figure 1. PRISMA flow diagram showing the phases of the systematic review.

(qualitative/quantitative). Numerical summaries were provided from these data. After screening all the documents, we conducted a full text exploration. The papers were assigned one or more keywords which were successively refined, narrowed, and in some cases combined or abandoned as we became more familiar with the dataset. Generative collaboration between the researchers working independently noted and numbered keywords before cross-referencing for comparability. This prompted further discussion, allowing the reworking and refining of these keywords until agreement was reached. It was a strongly iterative and comparative process of sorting and resorting ideas (Akerlind, 2002). Themes were determined from the keywords.

General characteristics of the research

In the following section, we present a summary of the general characteristics of the studies included in this review in terms of the publisher, location for the study, the scope and program level as well as the research design that was used to conduct the studies.

Publisher, location, scope and program level and research design

Thirty articles fulfilled the above criteria (see the Online Supplemental Material). Most of the studies were published in geography education journals (n = 19) such as the *Journal of Geography*, *International Research in Geographical and Environmental Education*, *Environmental Education Research* and *Teaching Geography*. The remainder of the articles (n = 11) were published in a variety of other journals including *International Journal of Learning, Teaching and Teacher Education*, *Curriculum Journal*, *Assessment in Education:*

Principles, Policy and Practice, *Issues in Educational Research*, *Cambridge Journal of Education*, *Learning and Individual Differences*, *International Journal of Progressive Education*, *New Educational Review*, and *Educational Research and Reviews*.

A total of 12 countries were represented with most of the studies conducted in the US (n = 6) and the UK (n = 6). Research had also taken place in five European countries including Bulgaria, Turkey, Germany, Spain and the Netherlands. In Africa, two countries were represented, namely Zimbabwe and Rwanda, with further studies from New Zealand (n = 1), Australia (n = 3) and one from India. What was noteworthy about the majority of these studies was that they took place at a national (n = 17) or local (n = 7) scale with only four crossing international boundaries. In terms of program level, most of the studies were conducted with secondary school students (n = 23) with only two undertaken in primary school settings. Two studies looked across primary and secondary contexts and three across secondary and tertiary.

The papers involved a range of designs including document analysis, experimental design, interviews and observations, policy analysis, think aloud protocol, case study and questionnaires. Broadly, the studies were classified as qualitative (53%), quantitative (18%) and mixed methods (29%) with participant numbers ranging from 10 to 2879 students/teachers. In the following section, we elaborate the results of the thematic analysis.

Themes and findings

Across the data set, we identified eight themes: (1) formative assessment; (2) spatial reasoning; (3) performance/achievement standards; (4) marking; (5) content analysis of assessment items (alignment with national policy requirements); (6) international assessment; (7) other assessment strategies (free word association, sketch mapping, self-evaluation); and (8) case studies and evaluations of national practice. Collectively these accounted for 30 papers.

Formative assessment

Five of the papers retrieved focused exclusively on Assessment for Learning (AfL) (or formative assessment) in geography. All of the studies were conducted in secondary schools in the Western world with the majority (n = 4) undertaken in the UK. The studies examined the role of feedback (Baldwin, 2003), university/school collaborations (Cooper & Cowie 2010), self-assessment (Davies, Durbin, Clarke, & Dale, 2004), and planning and implementation of AfL (Leat & Nichols, 2000; Tiknaz & Sutton, 2006). Three of the studies employed thematic analysis as the analytical approach, one was a narrative case study and the other used techniques associated with ethnography. Data collection sources ranged from exam scripts, interviews, classroom observations, recordings, reflections, surveys and meeting notes. It was noteworthy that none of these studies were conducted in primary school settings and only one used quantitative measures.

Collectively, the studies revealed that AfL can have a significant impact on student outcomes in geography. For example, Cooper and Cowie (2010) found that AfL enhanced student motivation and learner autonomy. Additionally, students with an understanding of assessment criteria could more effectively articulate quality and interpret teacher feedback to improve their scores (Davies et al., 2004; Leat & Nichols, 2000). It was recognised

that AfL practices were more sustainable with external support, shared teacher knowledge and beliefs, systematic professional experimentation, and reflexivity (Cooper & Cowie, 2010). Tiknaz and Sutton (2006) also highlighted the importance of statutory requirements and teachers' professional knowledge in the implementation of AfL in geography classrooms. Baldwin (2003), for example, found that even though teachers in his study (n = 6) were aware of the importance of formative assessment, they used few comments and generally did not use feed forward approaches aligned to achievement standards. Baldwin (2003) emphasised the need for policy documents in New Zealand to include exemplars of formative assessment practice.

Spatial reasoning

There were five papers related to geospatial thinking. Four took place in secondary or university settings and one in a primary context. Four of the five studies used a quantitative design involving the collection and statistical analysis of test data (Battersby, Golledge, & Marsh, 2006; Huynh & Sharpe, 2013; Lee & Bednarz, 2012; Tomaszewski, Vodacek, Parody, & Holt, 2015). These papers focused on progression of incidentally learned geospatial knowledge (Battersby et al., 2006) and the validation of an assessment instrument (STAT) (Lee & Bednardz, 2012; Tomaszewski et al., 2015). The other paper in this theme used a qualitative approach featuring document analysis to examine the extent to which spatial thinking concepts are being practiced in US schools (Anthamatten, 2010). All five papers were directly linked to North America; however, the final study involved the testing of an American instrument in Rwanda.

The earliest paper in this theme (Battersby et al., 2006) explored the progression of students' geospatial knowledge as they matured from middle-school through to university. Battersby et al. (2006) found a distinct hierarchy in terms of the acquisition of specific geospatial concepts. They concluded that a map overlay concept was too complex for the majority of middle school students to grasp without targeted instruction. For this reason, Battersby et al. (2006) argued for laying foundations for learning geospatial concepts early in the school curriculum. However, Anthamatten (2010) found that many key spatial thinking skills were absent from state social studies and geography standards documents. There was, therefore, a lack of guidance for teachers. He advocated for writers of standards to be cognisant of the research in this domain.

The three other studies in this theme examined the development of a standardised test for spatial thinking ability (Huynh & Sharpe, 2013; Lee & Bednarz, 2012; Tomaszewski et al., 2015). Lee and Bednarz (2012) developed the Spatial Thinking Ability Test (STAT) and demonstrated its reliability and validity in assessing the spatial thinking skills of 532 sary/university students in the US. Tomaszewski et al. (2015) refined the STAT instrument for the Rwandan cultural context and found that urban and male students outperformed rural and female students respectively in relation to spatial thinking skills. Huynh and Sharpe (2013) built on the work of Lee and Bednardz (2012) and Battersby et al. (2006) by developing a valid and reliable test of geospatial thinking that enabled educators to benchmark student performance and level of understanding as either novice, intermediate or expert. Their findings confirmed that there was a gradient of difficulty in the application of geospatial concepts and that geospatial thinking had multiple dimensions identified including analysis, comprehension, representation, application, scale and spatial relationship.

Performance/achievement standards

Two papers attempted to develop achievement standards for secondary geography students; however, each took a different approach. Whilst Davies (2002) used exam scripts to ascertain the level of sophistication of students' responses, ranking quality according to specificity, completeness and the ability to make a reasoned judgement, Lock and Barrett (2002) categorised learning outcomes in existing curriculum documents before proposing and validating a scale of achievement. Both mixed methods studies were conducted in secondary schools across multiple sites but at opposite ends of the globe (UK and Australia). Both papers advocated for the development and application of achievement standards in geography but posited that further research was required.

Marking

Two papers fell into this theme, namely Crisp (2008) and Panadero, Tapia, and Huertas (2012). While both studies were undertaken in multi-site secondary environments, they focused on different stakeholders in the marking process, examiners and students. Additionally, both studies used different methodologies. Crisp (2008) applied think-aloud protocols and statistical analysis to explore the psychological processes underlying the marking of AS1 and A2 Level geography examination scripts in the UK. The findings suggested that positive evaluations, comparisons and thorough reading were important to avoid marking severity and potential problems with reliability. On the other hand, Panadero et al. (2012) Spanish study focused on the relationship between student self-assessment (including the use of rubrics and sample scripts), self-regulation and self-efficacy. The findings indicated that self-assessment tools promoted higher levels of self-regulation with scripts outperforming rubrics. However, when it came to self-efficacy, the authors claimed that such self-assessment tools were inadequate for the promotion of mastery learning.

Content analysis of assessment items: alignment with national policy requirements

Another common theme in the empirical research was the alignment of assessment items with national policy including national standards, curriculum and examination specifications. Three studies all located in the national secondary contexts of China (Yang, 2013), Turkey (Yasar, 2009) and India (Mishra, 2015) explored the nature of assessment items in textbooks. The other two papers (Bijsterbosch, van der Schee, & Kuiper, 2017; Wertheim et al., 2013) within this theme focused on large-scale standardised tests and classroom assessments in the secondary context. All five papers investigated cognitive demand and knowledge type targeted.

The first two papers in this theme examined changes in textbook questions in response to national educational reforms. In the case of China, this referred to the implementation of the 2001 national standards, and in Turkey, changes in secondary education geography curriculum (2005). The Chinese study focused on question type, cognitive demand and political nature of items. The Turkish study examined whether assessment items reflected the curriculum reform; that is, a shift towards the assessment of learning processes rather than merely product. In general, both studies concluded that assessment items in textbooks had become more effective. For the Chinese, there was more flexibility in textbook

production with a greater emphasis on higher order thinking and a move towards less politicised questions. It was a similar story in Turkey with more diverse and better quality design assessment items promoting higher order thinking. The third paper in this theme, Mishra (2015), examined Indian geography textbook questions in relation to the National Curriculum Framework (NCF) using Jo and Bednarz's (2009) taxonomy of geospatial thinking. In contrast to the above studies, Mishra (2015) found that questions in Years 9 and 10 geography textbooks lacked a focus on authentic problem solving, did not assess higher order thinking skills, and reflected the regional approach to geography from the 1960s which was limited in scope.

Bijsterbosch et al. (2017) study revealed similar findings to that of Mishra (2015). However, the context was different. In their analysis of 39 internal examinations across 13 schools, they found that the majority of test items targeted lower levels of cognitive demand (mainly recall and to a lesser extent understanding) and conceptual and factual knowledge. Questions focused on rote learning rather than allowing students to demonstrate the application of geographical skills and knowledge of key models and theories. The authors cited a number of possible explanations for this including mirroring the external test to provide exam preparation (teaching to the test) and a lack of competence and confidence in writing and assessing inquiry based items. They also maintained that higher order cognitive processes are often the focus of AfL activities and were therefore not tested summatively.

Wertheim et al. (2013) examined the nature of assessment items in large-scale standardised tests and classroom assessments in K–12 US classrooms. As a part of the Road Map Project, they explored the alignment between current assessment practices and the criteria for effective assessment design derived from the geography content goals from *Geography for Life* and geography practices clarified in the *Road Map Assessment Report*. Consistent with the findings of Bijsterbosch et al. (2017) in the Netherlands, they concluded that assessment practices in the US focused primarily on the recall of factual and conceptual knowledge and that few items evaluated the ability of students to demonstrate geographic practices including evidence-based reasoning. The study also revealed widespread problems with item quality. Of the items studied, 60% were judged to have problems that would impede students' abilities to accurately represent what they know and what they can do with their geographical knowledge.

International assessment

In relation to international assessment, we identified three papers, namely van der Schee and Kolkman (2010), and Lane and Bourke (2016a, 2016b). The first paper described the nature of an existing cross-national assessment (The International Geography Olympiad) and analysed the results of the multimedia test using Pearson's correlation. This study found a strong correlation between the three parts of the 2008 Olympiad and the total score. The authors concluded that the International Geography Olympiad is a valid and reliable assessment that could be used to establish an international benchmark of geographical literacy.

The other two qualitative papers reported the findings of a survey of 74 international stakeholders from 36 countries and investigated: (1) the perceptions of the global geography education community on the advantages and challenges of initiating and

implementing a geography assessment in TIMSS; and (2) whether there could be a consensus in terms of what should be assessed and how the test should be implemented. In the first paper, the researchers found that the geography education community was overwhelmingly in favour of such a move because of the value of data obtained for research, policy and teaching practice. A number of questions about the development and implementation of this assessment were posed and explored in the second paper. The results indicated that the global geographical education community was divided in their views about the target age for an international assessment and the concepts and skills that should be targeted. The findings suggested that it was not possible to implement an international assessment of this kind without first establishing a common set of learning objectives for Years 4 and 8 that defined the types of knowledge and cognitive domains to be taught and assessed globally. The paper posited that this process was likely to result in some changes to local curricula to ensure there was a common framework for the assessment.

Other assessment strategies

This theme included an amalgam of methods for assessing geographical understanding, namely concept mapping, sketch mapping, argumentative writing and free word association. Two of the papers in this theme, Andrews, Tressler, and Mintzes (2008) and Wehry, Monroe-Ossi, Cobb, and Fountain (2012), both investigated concept mapping as a tool to assess conceptual understanding in geography. Andrews et al. (2008) paper focused on Years 6–9 students (n = 325) and adopted an experimental design, whereas Wehry et al. (2012) adopted a qualitative approach with Year 7 students (n = 43). Both were localised US studies which made similar conclusions about the value of concept maps as an alternative or adjunct to traditional pencil and paper tests. The studies revealed how the development of concept maps could provide valuable qualitative and quantitative information about students' depth of understanding (Andrews et al., 2008) and provided teachers with a resource for reflecting on their own practice (Wehry et al., 2012). Additionally, Wehry et al. (2012) argued that concept mapping techniques enhanced spatial thinking skills and stimulated interest in the subject under study, in this case, human geography.

Related research investigating visual representations as assessment tools in geography included the work of Harwood and Rawlings (2001). This study involved an examination of 26 primary students' freehand sketch maps of the world to assess their understanding of the spatial arrangement, size and shape of continents. The results highlighted the value of this approach for identifying misconceptions and blind spots in student understanding. Moreover, the authors argued that the process of constructing freehand sketches and receiving feedback from teachers and peers could improve primary students' world knowledge.

Various researchers have investigated the use of essays and argumentative writing as methods for assessing geographical understanding. Two papers in this theme included Budke, Schiefele, and Uhlenwinkel (2010) and Munowenyu (2007). Both studies were situated in secondary contexts but used different methods. Budke et al. (2010) German study counted the number and type of arguments used by students in one class. Texts were analysed first by argumentation diagramming (creating diagrams to show how parts of the argument relate to each other) and then by checking the structure of the arguments using

Toulmin's model. Munowenyu's (2007) study was larger, involving 83 students across three schools in Zimbabwe. Munowenyu applied the Structure of Observed Learning Outcome (SOLO) taxonomy and a marking schema developed by Lambert (1996) to determine the relative effectiveness of classroom versus fieldwork instruction. The author's analysis demonstrated the effectiveness of examining the structure and complexity of essay responses as a measure of understanding.

One study investigated methods for assessing students' preconceptions in geography. This cross-national study (Turkey and Bulgaria) by Atasoy (2009) with 174 Year 8 students used a free word association test to assess conceptions of key geographical and environmental terms. The author noted that associative thinking techniques were underutilised in geography instruction despite their effectiveness as both an assessment and learning tool for enhancing creative thinking.

Case studies and evaluations of national practice

The two papers in this theme, Pinar (2011) and Butt, Weeden, Chubb, and Srokoszc (2006) provided educators' views about assessment in their respective countries (Turkey and the UK). The qualitative Turkish study involving 10 teachers across multiple sites revealed that geography teachers were resistant to the introduction of problem solving and process based tasks which were central to the 2005 curriculum reforms. The participating teachers felt ill-equipped in terms of content knowledge and skills, and regularly reverted to their traditional familiar practices. The UK study was much broader and reported a snapshot of geography pedagogy and assessment across 17 state secondary schools in England providing a "window on practice and performance in assessment in geography" (Butt et al., 2006, p. 135). Amongst a range of findings regarding the General Certificate in Secondary Education (GCSE) geography specifications, curriculum planning and pedagogy, the authors concluded that geography teachers lacked a comprehensive understanding of the importance of formative assessment and target setting to maximise student learning outcomes. Both papers highlighted the need for key stakeholders to understand the rationale for educational reform and being professionally developed with the right knowledge and skills for implementation. This is consistent with the conclusions of Cooper and Cowie (2010) regarding the characteristics of sustainable reform (see Theme 1).

Discussion

The two key research questions for this study were: (1) what do we know about evidence-based assessment in geography education?, and (2) what gaps are identified to point to future research directions? The following discussion outlines the response to research question one and the conclusion will point to future research directions, thus answering research question two. In response to the first question, this review found 30 empirically based, peer-reviewed papers published in English over the past 16 years that focused on assessment in school-based geography education. Synthesis of the research suggests that we have some knowledge of assessment in geography education across the eight themes outlined above. Within these themes several distinctive trends were identified and are worthy of further elaboration.

For most, data collection periods in these geography education assessment studies were short with an absence of longitudinal research providing ongoing evidence of the impact of assessment on learning outcomes. Additionally, most were small scale with local and national studies (mainly UK and US based) accounting for 86% of the papers meeting the review criteria. Furthermore, there was a lack of connection between researchers exploring similar research questions in different places or contexts. For example, of the five papers included in the theme of assessment for learning, only one (Davies et al., 2004) cited any of the previous papers in this area of geography assessment. This created issues of transferability of the findings. So whilst there was research devoted to assessment in geographical education, there did not seem to be a coherent direction or collaboration on a global scale around a key set of research problems that needed to be investigated. It was also apparent that there was a lack of research around assessment in the primary sector.

The next trend apparent in the literature was a lack of research developing and validating instruments for assessing key geographical understandings and skills. The key exceptions were the work of Lee and Bednarz (2012) and Huynh and Sharpe (2013) who had developed and validated instruments for assessing spatial reasoning. These studies represented two of the limited number of quantitative papers exploring assessment in geography education. The Rwandan study by Tomaszewski et al. (2015) employed a mixed method approach. However, on the whole papers tended to be more qualitative in design.

This review also highlighted some alignment issues with assessment practices in geography education. The research suggested that while there had been a number of changes in educational standards and curriculum documents internationally, assessment practices had not always adjusted to align with these changes. For example, in the Indian study by Mishra (2015) there was a lack of alignment between the requirements of the national curriculum framework and the nature of geography textbook questions. This was consistent with research in Holland by Bijsterbosch et al. (2017) who found that internal assessments lacked a focus on complex knowledge and cognitive processes outlined in the syllabi. This affected the validity of assertions made from assessment data.

Another trend was the paucity of empirical studies examining progression in geographical understanding. The limited research in this area was acknowledged by the authors of the *Road Map for 21st Century Geography Education Project (Assessment)* who called for "a program of research [to] be initiated immediately to study learning progressions in geographic practices over the K–12 timeline" (Edelson et al., 2013, p. 66). This review only identified one empirical paper (Tiknaz & Sutton, 2006) on progression in geography understanding.

Conclusion

In response to research question two, there is a need for consensus regarding elements of geographical literacy and the development and validation of instruments for assessing such. Although the small scale studies can be informative in the local context, there is a need for large-scale, longitudinal studies with international researchers working collaboratively on an agreed set of research questions. The International Charter on Geographical Education (IGU-CGE, 2016) suggests that one of these questions should encompass how we understand students' learning progressions in geography and how these progressions are best calibrated and assessed. To enhance the validity of inferences from research, the

design of these studies should include qualitative, quantitative and mixed methods approaches. It is also clear that research on evidence-based assessment of geographical understanding in primary school contexts is required. Recent movements within the geography education landscape are addressing some of the issues identified above. For example, the National Center for Research in Geography Education (NCRGE) supported by the National Science Foundation in the US is facilitating research coordination, collaboration, and information sharing among geographers and educational researchers in other disciplines. Additionally, initial work has begun on the development of an international assessment of geographical literacy which will produce, for the first time, a valid and internationally accepted survey of young people's geographic knowledge, skills, and abilities. Data on such an international scale could be used to guide policy, practice and further research.

In summary, this systematic review confirmed the dearth of research on assessment in geography education and provided a structured approach that can be replicated for identifying where the gaps are. As a learning and research community, we need to know more about the essential knowledge and skills that students should develop, the nature of learning progressions associated with these knowledge and skills, and the types and formats of assessment instruments that will provide valid and reliable measures of this progress. This is the research agenda in assessment in geography education that needs to go forward.

Note

1. This report *reviews* current assessment frameworks and practices in K–12 geography education in the US. It then proposes a new approach to assessment in geography that will enable assessment developers to address the critical issues in assessment design. It is therefore considered a non-empirical paper for the purposes of this review. The primary research conducted as a part of the Road Map Project (e.g. Wertheim et al., 2013) is discussed under theme 5: Content analysis of assessment items: alignment with national policy requirements.

Disclosure statement

The authors declare that there are no conflicts of interest.

ORCID

Rod Lane http://orcid.org/0000-0002-7436-3424
Terri Bourke http://orcid.org/0000-0001-7298-9637

References

Akerlind, G. (2002, November). *Principles and practice in phenomenographic research*. Paper presented at the Current Issues in Phenomenography Conference, Australian National University, Canberra.

Andrews, K. E., Tressler, K. D., & Mintzes, J. J. (2008). Assessing environmental understanding: An application of the concept mapping strategy. *Environmental Education Research*, *14*(5), 519–536.

Anthamatten, P. (2010). Spatial thinking concepts in early grade-level geography standards. *Journal of Geography, 109*(5), 169–180.

Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: Treatments for myocardial infarction. *Journal of the American Medical Association, 268*(2), 240–248.

Atasoy, E. (2009). Eighth grade students' associations of geographical and environmental concepts. *New Educational Review, 17*(1), 102–124.

Atherton, R. (2011). Controlled assessment: Getting it right ? *Teaching Geography, 36*(1), 26–27.

Baldwin, R. (2003). Teachers' use of feedback comments to improve student achievement in standards-based assessment. *International Journal of Learning, 10*, 1197–1208.

Battersby, S. E., Golledge, R. G., & Marsh, M. J. (2006). Incidental learning of geospatial concepts across grade levels: Map overlay. *Journal of Geography, 105*(4), 139–146.

Bennetts, T. (2005a). Progression in geographical understanding. *International Research in Geographical and Environmental Education, 14*(2), 112–132.

Bennetts, T. (2005b). The links between understanding, progression and assessment in the secondary geography curriculum. *Geography, 90*(2), 152–170.

Bettis, N. C. (2001). Assessment issues in geographic education for the twenty-first century. *Journal of Geography, 100*(4), 172.

Bijsterbosch, E., van der Schee, J., & Kuiper, W. (2017). Meaningful learning and summative assessment in geography education: An analysis in secondary education in the Netherlands. *International Research in Geographical and Environmental Education, 26*(1), 17–35.

Budke, A., Schiefele, U., & Uhlenwinkel, A. (2010). 'I think it's stupid' is no argument: Investigating how students argue in writing. *Teaching Geography, 35*(2), 66–69.

Butt, G., Weeden, P., Chubb, S., & Srokoszc, A. (2006). The state of geography education in English secondary schools: An insight into practice and performance in assessment. *International Research in Geographical and Environmental Education, 15*(2), 134–148.

Cooper, B., & Cowie, B. (2010). Collaborative research for assessment for learning. *Teaching and Teacher Education, 26*(4), 979–986.

Crisp, V. (2008). Exploring the nature of examiner thinking during the process of examination marking. *Cambridge Journal of Education, 38*(2), 247–264.

Davies, P. (2002). Levels of attainment in geography. *Assessment in Education: Principles, Policy & Practice, 9*(2), 185–204.

Davies, P., Durbin, C., Clarke, J., & Dale, J. (2004). Developing students' conceptions of quality in geography. *Curriculum Journal, 15*(1), 19–34.

Edelson, D. C., Shavelson, R. J., & Wertheim, J. A. (2013). *A road map for 21st century geography education: Assessment*. Washington, DC: National Geographic Society.

Ford, J. D., & Pearce, T. (2010). What we know, do not know, and need to know about climate change vulnerability in the western Canadian Arctic: A systematic literature review. *Environmental Research Letters, 5*(1). 1–9. doi:10.1088/1748-9326/5/1/014008

George, J., Clarke, J., Davies, P., & Durbin, C. (2002). Helping students to get better at geographical writing. *Teaching Geography, 27*(4), 156–160.

Harwood, D., & Rawlings, K. (2001). Assessing young children's freehand sketch maps of the world. *International Research in Geographical and Environmental Education, 10*(1), 20–45.

Hofmann, M., Hinkel, J., & Wrobel, M. (2011). Classifying knowledge on climate change impacts, adaptation, and vulnerability in Europe for informing adaptation research and decision-making: A conceptual meta-analysis. *Global Environmental Change, 21*(3), 1106–1116.

Huynh, N. T., & Sharpe, B. (2013). An assessment instrument to measure geospatial thinking expertise. *Journal of Geography, 112*(1), 3–17.

International Geographical Union Commission on Geographical Education (IGU-CGE). (2016). *2016 International charter on geographical education*. Beijing: Author.

Jo, I., & Bednarz, S. (2009). Evaluating geography textbook questions from a spatial perspective: Using concepts of space, tools of representation, and cognitive processes to evaluate spatiality. *Journal of Geography, 108*(1), 4–13.

Lambert, D. (1996). Issues in assessment. In P. Bailey & P. Fox (Eds.), *Geography teacher's handbook* (pp. 187–201). Sheffield: Geographical Association.

Lambert, D. (2011). The lie of the land (revisited). *Teaching Geography, 36*(1), 24–25.
Lane, R., & Bourke, T. (2016a). Possibilities for an international assessment in geography. *International Research in Geographical and Environmental Education, 26*(1), 71–85. doi:10.1080/10382046.2016.1165920
Lane, R., & Bourke, T. (2016b). The inclusion of geography in TIMSS: Can consensus be reached ? *International Research in Geographical and Environmental Education, 26*(2), 166–176. doi:10.1080/10382046.2016.1178939
Leat, D., & Nichols, A. (2000). Brains on the table: Diagnostic and formative assessment through observation. *Assessment in Education: Principles, Policy & Practice, 7*(1), 103–121.
Lee, J., & Bednarz, R. (2012). Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography, 111*(1), 15–26.
Lock, G., & Barrett, R. (2002). Standards framework: Developing scales of achievement in post-compulsory education: A case study. *Issues in Educational Research, 12*(1), 35–48.
Marcello, J. S. (2009). A proposal for assessment in geography education. *Journal of Geography, 108*(4–5), 226–232.
Mishra, R. K. (2015). Mapping the knowledge topography: A critical appraisal of geography textbook questions. *International Research in Geographical and Environmental Education, 24*(2), 118–130.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*, 264–269.
Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M. ... PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1). 1–9. doi:10.1186/2046-4053-4-1
Munowenyu, E. (2007). Assessing the quality of essays using the SOLO taxonomy: Effects of field and classroom-based experiences by 'A' level geography students. *International Research in Geographical and Environmental Education, 16*(1), 21–43.
Murray, S., Nuttall, J., & Mitchell, J. (2008). Research into initial teacher education in Australia: A survey of the literature 1995–2004. *Teaching and Teacher Education, 24*(1), 225–239.
Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education, 14*(2), 149–170.
Oxman, A. D., & Guyatt, G. H. (1993). The science of reviewing research. *Annals of the New York Academy of Sciences, 703*(1), 125–134.
Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806–813.
Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell.
Pinar, A. (2011). Geography teachers' views on the assessment and evaluation instruments and methods used in the renewed geography curriculum. *Educational Research and Reviews, 6*(3), 334–341.
Plummer, R., de Loë, R., & Armitage, D. (2012). A systematic review of water vulnerability assessment instruments. *Water Resources Management, 26*, 4327–4346.
Rudel, T. K. (2008). Meta-analyses of case studies: A method for studying regional and global environmental change. *Global Environmental Change-Human and Policy Dimensions, 18*, 18–25.
Smothers, J. (2002). Geography standards: Instruction and assessment in action. *Geographical Education, 15*, 37–42.
Solem, M., Huynh, N. T., & Boehm, R. (Eds.). (2015). *Learning progressions for maps, geospatial technology, and spatial thinking: A research handbook*. Newcastle upon Tyne: Cambridge Scholars Publishing.
Tánago Gonzales, I., Urquijo, J., Blauhut, V., Villarroya, F., & De Stefano, L. (2016). Learning from experience: A systematic review of assessments of vulnerability to drought. *Natural Hazards, 80*(2), 951–973.

Tiknaz, Y., & Sutton, A. (2006). Exploring the role of assessment tasks to promote formative assessment in Key Stage 3 Geography: Evidence from twelve teachers. *Assessment in Education: Principles, Policy & Practice, 13*(3), 327–343.

Tomaszewski, B., Vodacek, A., Parody, R., & Holt, N. (2015). Spatial thinking ability assessment in Rwandan secondary schools: Baseline results. *Journal of Geography, 114*(2), 39–48.

van der Schee, J., & Kolkman, R. (2010). Multimedia tests and geographical education: The 2008 International Geography Olympiad. *International Research in Geographical and Environmental Education, 19*(4), 283–293.

Weeden, P. (2013). How do we link assessment to making progress in geography ? In D. Lambert & M. Jones (Eds.), *Debates in geography education* (pp.143–154). London: Routledge.

Weeden, P., & Lambert, D. (2006). *Geography inside the black box: Assessment for learning in the geography classroom*. London: Granada Learning.

Wehry, S., Monroe-Ossi, H., Cobb, S., & Fountain, C. (2012). Concept mapping strategies: Content, tools and assessment for human geography. *Journal of Geography, 111*(3), 83–92.

Wertheim, J. A., Edelson, D. C., Hildebrant, B., Hinde, E., Kenney, M., Kolvoord, R., … Shavelson, R. (2013). A road map for improving geography assessment. *Geography Teacher, 10*(1), 15–21.

Yang, D. (2013). Comparing assessments within junior geography textbooks used in mainland China. *Journal of Geography, 112*(2), 58–67.

Yasar, O. (2009). A comparative analysis of assessment and evaluation exercises included in geography textbooks written according to the 2005 secondary education geography curriculum and textbooks of the former curriculum in Turkey. *International Journal of Progressive Education, 5*(1), 45–68.