Lower secondary school students’ scientific literacy and their proficiency in identifying and appraising health claims in news media: a secondary analysis using large-scale survey data

Lena Victoria Nordheim, Kjell Sverre Pettersen, Birgitte Espehaug, Signe Agnes Flottorp, Øystein Guttersrud

ABSTRACT

Objectives Scientific literacy is assumed necessary for appraising the reliability of health claims. Using a national science achievement test, we explored whether students located at the lower quartile on the latent trait (scientific literacy) scale were likely to identify a health claim in a fictitious brief news report, and whether students located at or above the upper quartile were likely to additionally request information relevant for appraising that claim.

Design Secondary analysis of cross-sectional survey data.

Setting and participants 2229 Norwegian 10th grade students (50% females) from 97 randomly sampled lower secondary schools who performed the test during April-May 2013.

Outcome measures Using Rasch modelling, we linked item difficulty and student proficiency in science to locate the proficiencies associated with different percentiles on the latent trait scale. Estimates of students’ proficiency, the difficulty of identifying the claim and the difficulty of making at least one request for information to appraise the claim, were reported in logits.

Results Students who reached the lower quartile (located at −0.5 logits) on the scale were not likely to identify the health claim as their proficiency was below the difficulty estimate of that task (0.0 logits). Students who reached the upper quartile (located at 1.4 logits) were likely to identify the health claim but barely proficient at making one request for information (task difficulty located at 1.5 logits). Even those who performed at or above the 90th percentile typically made only one request for information, predominantly methodological aspects.

Conclusions When interpreting the skill to request relevant information as expressing students’ proficiency in critical appraisal of health claims, we found that only students with very high proficiency in science possessed that skill. There is a need for teachers, healthcare professionals and researchers to collaborate to create learning resources for developing these lifelong learning skills.

Strengths and limitations of this study

- The large and representative sample (n=2229) of lower secondary school students who responded to the science achievement test improves the external validity of our findings.
- Estimating students’ proficiencies and task difficulties using Rasch modelling, we could compare students’ proficiency in science with the difficulty of identifying and appraising a health claim in a fictitious brief news report.
- All achievement test items were piloted twice to ensure a valid and reliable measure of scientific literacy, and the use of a digitalised assessment tool reduced sources of errors.
- We did a secondary analysis of test data collected in 2013, thus a shift in proficiency in subsequent student cohorts may have occurred.
- Using raters to code responses to the open-constructed ‘news report’ item, there is a potential of miscategorising responses owing to rater subjectivity.

BACKGROUND

News media is a leading source of health and scientific information for the public, including adolescents and young people, who frequently encounter and share news and information through digital media. According to Eurostat, more than two-thirds of young people access online news media regularly. More than half also deliberately search for health information online, indicating health-related topics to be important for youth, especially for those aged 15 years and above.

Media reports of health research often address preliminary and poorly executed studies as sensational ‘breakthroughs’, leading to large discrepancies between the claims made and the underlying strength of the evidence. The result is confusing and
conflicting claims, for instance, about what to eat and drink to maintain good health—claims that influence peoples’ perceptions and actions of health. Knowledge about scientific methods and scientific concepts is assumed as a necessity for appraising the reliability of health claims. Health literacy initiatives at schools might help develop students’ skills in appraising claims, and some suggest that these skills may empower students to make informed decisions about health and well-being over the life course.13,14

Some claim that a minimum level of scientific literacy is a prerequisite for developing health literacy.13,15 The aim of compulsory science education is to develop students’ scientific literacy, including the proficiency to design and evaluate scientific inquiry, and gain knowledge about how the procedures of science support or disprove claims.16 School science may therefore be a key learning area for developing adolescents’ proficiency to critically appraise health claims in the media. Importantly, educational frameworks promote media reports of research as important real-life contexts for advancing and assessing students’ scientific literacy in terms of evidence appraisal.16–18

Without appropriate training, adolescents find it difficult to engage critically with media reports containing scientific content, and this challenge continues as they move from compulsory to higher education.19–27 Studies indicate that students tend to overestimate the certainty of scientific claims and accept them at face value.19–21 23 26–27 Moreover, they rely on substitute credibility indicators such as expertise (eg, researchers, journalists) and authors’ use of scientific statements and prompts (eg, ‘evidence-based’ or ‘scientifically proven’) without any in-depth conceptual understanding.19 20 28

The majority of these studies reside within the body of research on scientific literacy, not health literacy. This reflects that critical thinking around science-related claims in media, including the proficiency to appraise the science behind health claims, are underscored themes in models and definitions of health literacy.29 30 Accordingly, these issues are hardly emphasised in measures and empirical studies of adolescents’ and young people’s health literacy.31–34

There has been a call for studies that explore how people’s scientific literacy correspond to their proficiency in accomplishing specific tasks associated with their health literacy, such as identifying and appraising health claims (National Academies of Sciences, p. 107 [35]). A relevant question concerns ‘what someone who scores in the upper quartile on a science literacy measure can do that someone who scores in the lowest quartile cannot?’ (National Academies of Sciences, p. 107 [35]). Our study aims to address this question, using data from a national science achievement test of Norwegian 10th grade students. We explore responses to an item designed as a brief news report of a fictitious scientific study that assessed students’ proficiency to identify and appraise a health claim.

In Norway, grade 10 is the final year of compulsory education and most students are 15 years of age. According to the Programme for International Student Assessment (PISA) studies of those aged 15 years, Norwegian students perform slightly above the OECD average in science (OECD, p. 44 [36]) and approximately 80% and 30% perform at or above PISA proficiency level 2 and 4 in science, respectively (OECD, p. 320 [36]). At level 2, students can typically ‘use common scientific knowledge to identify a valid conclusion from a simple data set’ (OECD, p. 68 [36]) and hence identify scientific claims—a prerequisite for appraising claims.20 21 37 At level 4, students can typically ‘identify the evidence supporting a scientific claim’ and draw on knowledge about scientific procedures (eg, experimental designs) to justify conclusions (OECD, pp. 72–4 [36]). Hence, they can most likely request further evidence when encountering unsupported science-based claims, a hallmark of critical appraisal.38

Previous studies suggest that students, if they request information, usually emphasise methodological aspects of the reported research, the findings as such, and theoretical explanations of the findings.24 25 27 37 39 40

Building on knowledge from prior research and applying the national science achievement test of Norwegian 10th grade students as a measure of scientific literacy, we hypothesised that:

1. Students who score at or above the lower quartile on the scientific literacy measure are proficient in identifying a health claim among other competing textual information.
2. Students who score at or above the upper quartile on the scientific literacy measure are proficient in both identifying a health claim and formulate at least one request for further information relevant for appraising that claim, predominately information about either the research methods applied, the data collected or the underlying mechanisms causing an outcome.

METHODS

Design

We did a secondary analysis of existing data from a large-scale cross-sectional, web-based science achievement test assessing a random sample of the 2013 cohort of 10th grade students in Norway.

Participants

In 2013, the cohort of 10th grade students comprised about 64,000 individuals distributed across 1238 schools.41 Using random sampling, excluding special schools and international schools, 200 public schools were contacted for consensus of participating in the voluntary student assessment. Eligible schools were selected with a probability-proportional-to-size sampling. No schools selected themselves into the study. All schools were contacted by email and telephone between 20 December 2012 and 6 February 2013. One class at each of 97 schools—a total of 2229 students (50% females), completed the digitalised
assessment during April–May 2013. We estimated the school/class average participation rate as 86%. Owing to technical shortcomings beyond our control, no data on students’ socioeconomic status or ethnicity was recorded. The mean final assessment grade in science at each school was available from the Norwegian Directorate for Education and Training. On a scale from 1 to 6, where 6 is best, the sample average grade in science was 4.0—identical to the eligible population average. No experimental manipulations or interventions were part of our study.

**Participant and public involvement**

Participants were not involved in the development of any part of this study.

**Study context**

In Norway, the integrated subject ‘natural science’ is a mandatory subject throughout compulsory education. At the time of the survey (spring 2013), the natural science curriculum was structured into six subject domains: ‘body and health’, ‘diversity in nature’, ‘the universe’, ‘phenomena and substances’, ‘technology and design’ and ‘the budding researcher’.

The latter is a cross-cutting domain to ensure that knowledge about science as a process is integrated more systematically throughout science domains.

The national science achievement test assessed students’ proficiency in science based on the competence aims in the science curriculum for grade 8–10, with assessment items distributed across the cognitive domains ‘knowing’ (knowledge of scientific facts, concepts and procedures), ‘applying’ (apply knowledge to explain phenomena and solve problems) and ‘reasoning’ (evaluating scientific enquiry and alike). The items were distributed across the science domains and cognitive domains as described in online supplementary file 1, the 2013 assessment emphasised the science domains ‘body and health’ and ‘diversity in nature’.

**Test items and the administration procedures**

The 54 test items constituted a sufficiently valid and reliable scale for measuring scientific literacy as defined by the Norwegian curriculum. All but the one open-constructed news item, positioned at the end of the assessment test and scored 0–4 points, were dichotomously scored selected-response items. Accordingly, the science test data were analysed against the partial credit parameterisation of the unidimensional Rasch model. By sampling items from a bank of prior field-tested items, it was possible to construct a scale with difficulty well-targeted at the population of interest. The ‘test reliability’ was acceptable (Cronbach’s alpha based on completely scored data=0.93; person separation index based on person proficiency estimates=0.92). Measured up against the applied Rasch model all but one item discriminated sufficiently well between students with low and high standing on the latent trait (scientific literacy), and no significant differential item functioning, violations of unidimensionality or local independence were observed. The one poorly discriminating item was discarded and the analysis was re-run. The students completed the test within 80 min at school using a digitalised assessment tool.

The open-constructed ‘news item’ was designed to evaluate students’ proficiency in identifying and critically appraising a health claim. The item’s stem was designed as a brief news report (70 words) that referred to a fictitious study concluding that eating corn regularly reduces the risk of type II diabetes. The content and format of the item was similar to the news brief items in an instrument developed by Korpan et al, with no details about the study except being conducted by American scientists. In addition, there was a brief background statement about the rising global prevalence of type II diabetes along with a declaration from a diabetes interest group promoting the study findings. Students were first asked to identify the health claim in the news report, more specifically the conclusion from the fictitious study (the word ‘conclusion’ was used in the item’s question), that is, a regular intake of corn reduces the risk of type II diabetes. Second, they were asked to generate requests for information about the study that they would need to appraise the reliability of the health claim. Students were instructed to write a maximum of one and two sentences for the health claim and requests, respectively. A 250-character limit on students’ responses was imposed by the electronic assessment system (beyond our control). Responses to the news item allowed us to assess aspects of students’ functional and critical health literacy, more specifically their comprehension of health information and claims, and their ability to critically appraise claims. The item has been retained for continuous test use and is thus unavailable for publication.

**Analysis**

We coded responses to the news item using a coding guide of assessment criteria that reflected both credited and non-credited responses with regard to identifying the health claim (first part of the item) and requesting information about the study referred to in the item’s stem (second part of the item). The process of coding students’ information requests was based on a taxonomy for classifying questions and knowledge about scientific research. See table 1 for an overview of the coding guide, including the taxonomy’s main scientific research categories (eg, methods). We continually improved the coding guide during field tests and clarified it by including examples of authentic student responses (see online supplementary file 2 for a complete version of the guide).

One rater coded all student responses and consistency was evaluated by using an additional rater who coded 25% of the responses. Inter-rater agreement (OG and KSP) for the health claim was 94% and improved to 96% after discussion. Inter-rater agreement (LVN and KSP) for the information requests was 86% and improved to 98% after discussion. The lower initial agreement rate
Table 1 Overview of coding guide for the news item

<table>
<thead>
<tr>
<th>Part 1: health claim</th>
<th>Credited response</th>
<th>Non-credited response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A response…</td>
<td>A response…</td>
</tr>
<tr>
<td>Complete</td>
<td>providing a complete account of the claim (ie, that a regular intake of corn reduces the risk of type II diabetes).</td>
<td>where the claim relates to the topic of the news report, but is otherwise wrong.</td>
</tr>
<tr>
<td>Mostly complete</td>
<td>providing a mostly complete account of the claim, with some significant words lacking (ie, ‘regular’ and/or ‘type II’), and/or referring to amount of intake (eg, ‘much’ or ‘more’ corn).</td>
<td>with no reference to corn and/or type II diabetes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 2: information requests</th>
<th>Credited response</th>
<th>Non-credited response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A response relating to…</td>
<td>A response…</td>
</tr>
<tr>
<td>Methods</td>
<td>how the study was conducted, including study design, subjects, procedures and measurements.</td>
<td>indicating the need for one or more future studies, either in general, or relating specifically to one of the scientific research categories (methods, data, etc).</td>
</tr>
<tr>
<td>Data/statistics</td>
<td>what was observed in the reported study, or about statistical tests used to analyse the data.</td>
<td>indicating that the student does not believe that the study has been conducted.</td>
</tr>
<tr>
<td>Theory/agent</td>
<td>why the reported effects might have occurred, including questions about the properties of the presumed causal agent and/or possible underlying mechanisms.</td>
<td>relating to the topic of the news report, but is otherwise wrong.</td>
</tr>
<tr>
<td>Social context</td>
<td>the credentials and bias related to who did the study or funded it and where it was conducted or published.</td>
<td>only vaguely referring to the scientific categories (methods, data, etc).</td>
</tr>
<tr>
<td>Relevance</td>
<td>the importance or applicability of the study findings, or the impact of the study.</td>
<td>which is irrelevant or is a ‘do not know’ response.</td>
</tr>
<tr>
<td>Related research</td>
<td>whether the findings have been replicated or fit results from previous research.</td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td>the study described in the news report that is ambiguous because it fits under two or more scientific categories.</td>
<td></td>
</tr>
</tbody>
</table>

When coding part 1, the raters applied one variable and used values starting with ‘1’ and ‘0’ to indicate whether the response included an acceptable account of the health claim or not. For part 2, the raters applied eight variables. Seven of these were labelled to reflect the scientific research categories (methods, etc), raters used the values ‘1’ and ‘0’ to indicate whether the response included an acceptable request for information within the specific category. For the eight variable, non-credited responses to part 2, values starting with ‘0’ were used to indicate type of response. Blank responses (part 1 or part 2) were coded with the value ‘99’.

was mainly owing to interpretation of responses that concerned the need for future studies (see specifications in table 1). As the item’s stem explicitly asked students to relate their requests to the specific study presented in the news report, our final decision was not to credit responses that concerned ‘future studies’.

Overall, we credited students’ responses to the news item according to a ‘full credit’ (4 points), ‘partial credit’ and ‘no credit’ (0 point) system as specified in the scoring guide (table 2). This cumulative scoring guide made it possible to identify a student’s skill simply by knowing that student’s item score. We considered it unlikely that students who failed to identify the health claim were able to request information needed to establish the reliability of that claim. Thus, an acceptable account of the claim, as specified in table 1, was a premise for being credited on the item.

The software package RUMM2030 was used for Rasch modelling. Using unidimensional Rasch modelling, one may construct a scale and locate each item’s threshold(s) on that scale. A dichotomously scored item has one threshold reflecting the difficulty of the item, and a polytomously scored item has $k-1$ thresholds reflecting the difficulty of its $k$ score categories. The news item had...
### Table 2: News item scoring guide

<table>
<thead>
<tr>
<th>Credit</th>
<th>Score categories</th>
<th>Type of response</th>
</tr>
</thead>
<tbody>
<tr>
<td>No credit</td>
<td>0</td>
<td>Wrong or vague health claim, irrelevant or blank response.</td>
</tr>
<tr>
<td>Partial credit</td>
<td>1</td>
<td>Acceptable account of the health claim.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Acceptable account of the health claim and requests relating to one scientific research category (eg, methods).</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Acceptable account of the health claim and requests relating to two unique scientific research categories (eg, methods and data).</td>
</tr>
<tr>
<td>Full credit</td>
<td>4*</td>
<td>Acceptable account of the health claim and requests relating to three or more unique scientific research categories (eg, methods, data and theory/agent).</td>
</tr>
</tbody>
</table>

*No student made more than three unique requests for scientific information. Accordingly, full credit on the item was set to 4 score points.

---

**RESULTS**

Two-thirds (64%) of the students identified the health claim, of whom only half gave a complete account of it (table 3). Figure 1 shows that the difficulty associated with identifying the health claim was 0.0 logits (score point 1), that is, it equals the mean difficulty of the test items in the national achievement test, which was set to 0.0. Accordingly, the average scientific literate student was likely able to identify the health claim in the brief news story. Students who reached the lower quartile on the scientific
By coding and analysing these responses, we concluded that the technical deficit might have constrained the limit and were thus truncated by the assessment system.

We assessed how 10th grade students’ levels of scientific literacy corresponded to their proficiency in identifying and critically appraising health claims in the news media—two essential aspects of health literacy. The findings weakened our first hypothesis, as only the average proficiency estimate for this group was 0.0 logits, which equals the difficulty of identifying the health claim. Students who made suggestions for the conduct of future studies, rather than making requests for information related to the study reported in the news story (and thus were not credited), performed somewhat better on the achievement test (average proficiency 0.64 logits).

Characteristics of students’ information requests
As shown in table 3, and in line with hypothesis 2, the most frequent requests were related to how the study had been conducted (methods), the data collected (data/statistics) and the theoretical explanations of the results (theory/agent). The requests across these topics varied in level of detail (table 4). More than half of the requests about data/statistics were rudimentary. In comparison, all requests about theory/agent were specific, for example, concerning what active ingredient in corn actually caused the preventive effect. Methods was the only topic where several students made more than one request about specific features of the topic. Nearly half of these requests concerned the study participants, primarily the sample size (126 of 230 requests). Less frequent were requests about design, including the control of confounding variables and use of control groups (33 of 60 requests). As these requests belonged to the same unique scientific research category (methods), they were credited only once (table 2).

Seventy-one per cent of students provided a response that was either blank or otherwise disapproved, and were thus assigned a ‘non-credit’ category. The average proficiency estimate for this group was 0.0 logits, which equals the difficulty of identifying the health claim. Students who made suggestions for the conduct of future studies, rather than making requests for information related to the study reported in the news story (and thus were not credited), performed somewhat better on the achievement test (average proficiency 0.64 logits).

DISCUSSION
We assessed how 10th grade students’ levels of scientific literacy corresponded to their proficiency in identifying and critically appraising health claims in the news media—two essential aspects of health literacy. The findings weakened our first hypothesis, as only the average scientific literacy level of students aged 15 years leaving compulsory school was able to identify a clearly stated health claim in a rather simple news report. Students performing at the upper quartile of the scientific literacy measure were barely proficient at identifying and appraising the claim, namely making a request for evidence needed to determine the reliability of the claim. Accordingly, our second hypothesis was strengthened.

About half of the invited schools participated, and the average student participation rate at these schools was at an acceptable 86%. Our analyses indicated that the sample school average grade and gender distribution matched the population distributions at grade 10, thus indicating generalisability of our findings. Although data on socioeconomic status and ethnicity was unavailable for this study, previous studies have found these factors to predict science proficiency in Norwegian students aged literacy scale were not likely to identify the claim, as their proficiency (-0.5 logits) was 0.5 logits below the difficulty estimate of score point 1. Hence, hypothesis 1 was weakened as students’ skills were much poorer than we expected based on our interpretation of PISA results.

Less than one-third of the students (29%) made one or more information requests about the reported study relevant to appraise the health claim (table 3). Figure 1 indicates that the difficulty associated with score point 2 (1.5 logits)—identifying the claim and making one request for information, was rather close to the proficiency associated with the upper quartile (1.4 logits). Therefore, hypothesis 2 was strengthened. However, students located even at or below the 90th percentile (2 logits) were not likely to score ≥2 points on the news item, that is, identifying the health claim and making more than one request for information.

A few responses (n=115) exceeded the 250-character limit and were thus truncated by the assessment system. By coding and analysing these responses, we concluded that the technical deficit might have constrained the opportunities to achieve a higher score (ie, to be credited for further information requests) for 31 students only.

Table 3 Proportion of students who identified the health claim of the (news item part 1) and the proportions who requested different types of information for appraising that claim (part 2)

<table>
<thead>
<tr>
<th>N (of 2229)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part 1: health claim</strong></td>
<td></td>
</tr>
<tr>
<td>Credited responses</td>
<td>1420</td>
</tr>
<tr>
<td>Complete</td>
<td>710</td>
</tr>
<tr>
<td>Mostly complete</td>
<td>710</td>
</tr>
<tr>
<td>Non-credited responses</td>
<td>809</td>
</tr>
<tr>
<td>Wrong, vague or other</td>
<td>415</td>
</tr>
<tr>
<td>Blank</td>
<td>394</td>
</tr>
<tr>
<td><strong>Part 2: information requests</strong></td>
<td></td>
</tr>
<tr>
<td>Credited responses</td>
<td>652*†</td>
</tr>
<tr>
<td>Methods</td>
<td>376</td>
</tr>
<tr>
<td>Data/statistics</td>
<td>189</td>
</tr>
<tr>
<td>Theory/agent</td>
<td>146</td>
</tr>
<tr>
<td>Social context</td>
<td>57</td>
</tr>
<tr>
<td>Relevance</td>
<td>12</td>
</tr>
<tr>
<td>Related research</td>
<td>9</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>31</td>
</tr>
<tr>
<td>Non-credited responses</td>
<td>1577</td>
</tr>
<tr>
<td>Future studies</td>
<td>365</td>
</tr>
<tr>
<td>Disbelief</td>
<td>79</td>
</tr>
<tr>
<td>Wrong, vague or other</td>
<td>618</td>
</tr>
<tr>
<td>Blank</td>
<td>515</td>
</tr>
</tbody>
</table>

*Comprise all students who made one or more information requests, including the 50 students who were not credited on the news item part 1 (health claim).
†The total sum of requests will exceed 652 as 154 students made requests relating to more than one unique scientific category.
Table 4  Characteristics of the requests related to methods, data and theory (specifications of table 3)

<table>
<thead>
<tr>
<th>Type of request</th>
<th>Examples of students' responses</th>
<th>No. of requests</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rudimentary</td>
<td>How was the study conducted?</td>
<td>471</td>
<td>100</td>
</tr>
<tr>
<td>Design</td>
<td>Did they use a control group?</td>
<td>80</td>
<td>17</td>
</tr>
<tr>
<td>Agent delivery</td>
<td>How much corn is necessary to</td>
<td>60</td>
<td>13</td>
</tr>
<tr>
<td>(procedure)</td>
<td>eat? How often?</td>
<td>100</td>
<td>21</td>
</tr>
<tr>
<td>Participants</td>
<td>How many were tested?</td>
<td>230</td>
<td>49</td>
</tr>
<tr>
<td>Measures</td>
<td>Did they measure the participants' blood sugar?</td>
<td>1</td>
<td>&gt;1</td>
</tr>
<tr>
<td><strong>Data/statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rudimentary</td>
<td>We need the results.</td>
<td>191</td>
<td>100</td>
</tr>
<tr>
<td>Absolute nature of data*</td>
<td>For how many did this (not) work?</td>
<td>111</td>
<td>58</td>
</tr>
<tr>
<td>Comparative nature of data*</td>
<td>What is the reduction in risk? How large was the effect?</td>
<td>32</td>
<td>17</td>
</tr>
<tr>
<td>Duration of effect</td>
<td>Does it work over time?</td>
<td>43</td>
<td>23</td>
</tr>
<tr>
<td><strong>Theory/agent</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td>We need to know what corn is composed of (nutrients).</td>
<td>149</td>
<td>100</td>
</tr>
<tr>
<td>Agent mechanisms</td>
<td>What ingredient in corn prevents type II diabetes? Why does corn prevent type II diabetes?</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(Side) effects</td>
<td>Is there any side effect of eating corn?</td>
<td>117</td>
<td>79</td>
</tr>
<tr>
<td>Alternative agents</td>
<td>Do other corn products have this effect?</td>
<td>23</td>
<td>15</td>
</tr>
</tbody>
</table>

*For the dependent variable (here: diabetes risk).

15 years. For socioeconomic status, however, the relationship with proficiency is relatively weak compared with most other countries.

Previous studies of students’ evaluations of scientific claims have mostly been conducted at upper secondary school level and above, and on smaller, mostly self-selected samples of students. Thus, to our knowledge, this is the first study investigating students’ critical appraisal of science-based health claims in the context of a large student sample at lower secondary school level. Furthermore, while students’ scientific health knowledge in important areas such as chronic and infectious diseases and their knowledge of sources to science-based health information has previously been explored, we have addressed a call for research on the utility of scientific literacy for critical appraisal of health claims. The analysis of responses to the news item provided useful information about how proficient students were (ie, the levels of scientific literacy necessary to involve in identifying and appraising health claims in news reports of science), and what kind of knowledge they possessed and applied when approaching such claims (ie, which responses earned credits and which did not). Such in-depth knowledge of students’ thinking around a topic or task is an important outcome of secondary analyses of individual test items used in large-scale surveys, as it may formatively inform and develop teachers’ practices.

There were some limitations to this study. First, the test data were collected 6 years ago (2013), thus we acknowledge that a possible shift in students’ knowledge might have occurred. Nevertheless, our study is timely due to a major revision of the curriculum that will be implemented from autumn 2020. Second, to avoid response dependence between similar items, and accordingly violations of local independence in the data, the test comprised only one of the news item developed and field-tested. This prohibited us from evaluating whether variations in text dimensions (eg, the claim’s plausibility and how familiar the students were with the health topic) could have influenced students’ information request—dimensions previously reported to impact on students’ critical engagement with news reports. For the same reason, and because the news item did not include any embedded attitudinal items, we were unable to assess whether important personal factors (eg, interest in the health topic, belief in the claim, scientific attitude) could have affected students’ requests. Third, to make students respond shortly and ‘on task’, they were encouraged to write only two brief sentences. This might have constrained their opportunities to make several requests for evidence regarding the claim, although our analysis of incomplete responses due to the limit of 250 characters indicated this was probably not the case. Moreover, our findings resemble previous studies with regard to both the number and type of requests made.

Finally, while it is common practice to retain test items for re-use, this implies a lack of transparency in the test data used for this study.

Being able to correctly identify the nature of information included in media reports of science is a
prerequisite for critical appraisal. High school and university students exposed to news reports containing a variety of scientific features, often confuse conclusion statements with statements about the results (data) and explanations (theory).20 21 37 In comparison, the news item in our study was less complex, including only a few statements beside the health claim (study’s conclusion). Still, only students being proficient at or above the average in science managed to identify the claim, often providing an incomplete account of it. Therefore, the underlying problem seems to be the same as noted in previous studies, namely a lack of training in reading media reports of science.20 21 37 As previously noted, we were unable to explore whether the students perceived the claim as plausible or not. The former might be the case as uncertainty about a claim’s plausibility has been found to provoke more methods questions.37 39 For instance, students’ occasional requests about procedures indicate that they believed in the claim, and thus simply wanted information about how much, how often and how long the intake of corn should be to see the reported effect on diabetes risk. A further observation was the few requests about the social context of the research (eg, the American scientists’ affiliation), perhaps suggesting that students regarded science and scientists as authoritative and accordingly that the claim was plausible. This has also been noted in previous studies of students across educational levels.19–21 23 25–27

Our findings are of concern as they illustrate students’ functional and critical health literacy at the end of compulsory school. Almost three-quarters of the 10th grade students were unable to identify and critically appraise the health claim in a brief news report. Even the highest performing students mostly requested only one scientific research category of an optimum of six broad categories. Despite curricular mandates to develop scientific literacy and critical appraisal skills important for health literacy, it seems like students’ actual skills are underdeveloped and not taught in a way that improve their appraisal of health claims as assessed by the news item. This is consistent with findings from a qualitative study where science teachers reported that opportunities for teaching critical appraisal during inquiry-based activities, such as online health information seeking or small-scale experiments, are lost in the need to emphasise factual knowledge on health topics.54 Importantly, teachers do not acknowledge the relevance of teaching critical appraisal, or they lack methods to teach them. Thus, there is clearly a prospect for cross-sector collaboration between healthcare and education professionals and researchers to work together to enable laypersons to appraise health claims, as pointed out by Sharples et al.14 Research is still scarce as to which interventions best improve students’ ability to appraise health claims.55 56 However, a recent cluster-randomised controlled trial shows promising effects of a cross-disciplinary developed intervention aimed at teaching primary school children to appraise claims about treatment effects.57

Our study has identified specific areas that require attention in further development and evaluation of interventions—areas that align with important key concepts lay people need to know to assess health claims.11 It was encouraging that the students in our study—when they employed scientific criteria—were sensitive to methodological information, which often is lacking in media reports of health research.7 However, students requested only a limited range of methodological evidence, with little attendance to details about the study design, such as the use of control groups or control of confounding variables. This was noteworthy given the news report’s assertiveness in claiming a causal relationship between the intake of corn and the reduced risk of type II diabetes, mirroring the many misleading media reports that fail to differentiate association from causation.67 Science instruction should therefore develop students’ knowledge of good and weak designs for establishing a cause-effect relationship, including the design of controlled studies, the importance of fair comparisons, the principles of randomisation and blinding and proper and improper ways of reporting outcomes (eg, absolute vs relative risk). Existing evidence suggests that such knowledge is better gained through teacher-guided investigations that allow students to reflect on adequate and adequate experimental strategies, rather than through student-led hands-on or virtual experiments.38 Importantly, teachers need to make explicit the link between experiments, critical reading, and appraisal of health claims in the new(s) media. In our study, several students suggested the conduct of a future study rather than requesting information about the reported study, often involving themselves in doing so (eg, “we have to test a number of people”). This perhaps supports the notion that teachers enforce experimentation and hands-on activities without linking relevant learning outcomes to reading and critically appraising science presented in out-of-school contexts, including media reports.39 40 Finally, students hardly requested related research supporting or disproving the claim. Accordingly, teaching could sensitize students to the limitations of single studies, introducing the idea of systematic reviews.

Author affiliations
1Centre for Evidence-Based Practice, Western Norway University of Applied Sciences, Bergen, Norway
2Faculty of Medicine and Dentistry, Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway
3Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway
4Norwegian Institute of Public Health, Oslo, Norway
5Institute of Health and Society, University of Oslo, Oslo, Norway
6Faculty of Mathematics and Natural Sciences, Norwegian Centre for Science Education, University of Oslo, Oslo, Norway

Acknowledgements The authors would like to thank Joråun Østerholt Dalane who, together with KSP, developed the original version of the news brief item. The authors would also like to thank Bjørn Vidnes, Anders Isnes and Kirsten Fiskum who, together with ØG and LVN, developed all the other science achievement test items.

Contributors LVN, KSP and ØG developed a revised version of the news brief item and the coding guide, and coded the student responses. ØG administered the survey, managed the data handling, conducted the Rasch analysis and constructed the achievement scale. LVN wrote the preliminary draft of the paper.
with contributions from ØG. The paper was critically reviewed by KSP, BE and SAF for important intellectual content.

**Funding**
The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** The data set used in this paper did not include information that identified individuals, so ethical approval was not required under Norwegian regulations.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** No additional information is available. A data file with a sample of the coded student responses (in Norwegian) is available on reasonable request. Please contact Øystein Gutttersrud (oystein.guttersrud@naturfagsenteret.no).

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

**ORCID iD**
Lena Victoria Nordheim http://orcid.org/0000-0001-7370-1121

**REFERENCES**


5. Park E, Kwon M. Health-related Internet use by children and adolescents: systematic review. *J Med Internet Res* 2018;20:e120.


47. RUMM Laboratory Pty Ltd. RUMM2030 [program] 2011.


