



BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## Lower secondary school students' scientific literacy and their proficiency in identifying and appraising health claims in news media: a large-scale cross-sectional survey

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-028781
Article Type:	Research
Date Submitted by the Author:	22-Dec-2018
Complete List of Authors:	Nordheim, Lena Victoria; Western Norway University of Applied Sciences, Centre for Evidence-Based Practice Pettersen, Kjell Sverre; Oslo Metropolitan University, Faculty of Health Sciences Espehaug, Birgitte ; Western Norway University of Applied Sciences, Centre for Evidence-Based Practice Flottorp, Signe; Norwegian Institute of Public Health; University of Oslo, Institute of Health and Society Guttersrud, Øystein; University of Oslo, Faculty of Mathematics and Natural Sciences, Norwegian Centre for Science Education; Oslo Metropolitan University, Faculty of Health Sciences
Keywords:	Community child health < PAEDIATRICS, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

SCHOLARONE™  
Manuscripts

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

Lena V. Nordheim<sup>1\*</sup>, Kjell Sverre Pettersen<sup>2</sup>, Birgitte Espehaug<sup>1</sup>, Signe Flottorp<sup>3,4</sup>, Øystein Guttersrud<sup>2,5</sup>

<sup>1</sup>Centre for Evidence-Based Practice, Western Norway University of Applied Sciences, Bergen, Norway.

<sup>2</sup>Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway.

<sup>3</sup>Norwegian Institute of Public Health, Oslo, Norway.

<sup>4</sup>Institute of Health and Society, University of Oslo, Oslo, Norway

<sup>5</sup>Faculty of Mathematics and Natural Sciences, Norwegian Centre for Science Education, University of Oslo, Oslo, Norway

\*Corresponding author: Lena V. Nordheim, Centre for Evidence-Based Practice, Western Norway University of Applied Sciences, Bergen, Norway.

Inndalsveien 28

5020 Bergen

Norway

E-mail: [lvn@hvl.no](mailto:lvn@hvl.no)

Word count: 3870

## ABSTRACT

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

**Design:** Cross-sectional survey.

**Setting:** This study used data from a national science achievement test of Norwegian 10th grade students.

**Participants:** 2229 students (50% females) from 97 randomly sampled schools who performed the achievement test during April-May 2013. Special schools and international schools were excluded.

**Primary and secondary outcome measures:** The proficiency of students located at different percentiles on the latent trait scale, the difficulty of identifying the claim, and the difficulty of making at least one request for information. Using Rasch analysis, the proficiency and difficulty estimates are 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 very likely to identify the health claim but barely proficient at making one request for information (task located at 1.5 logits). Even those who performed at or above the 90th percentile mostly 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 possess that expertise. There is a need for teachers, healthcare professionals and researchers to collaborate to create learning resources for developing these lifelong learning skills.

ARTICLE SUMMARY

Strengths and limitations of this study

- This study included a large and representative sample (n = 2229) of Norwegian lower secondary students who participated in a national science achievement test.
- Estimating students’ proficiencies and task difficulties using Rasch analysis, we could compare students’ proficiency in science to 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 utilizing a digitalized assessment system reduced sources of errors.
- Using raters to code responses to the open-constructed “news report” item, there is a potential of misclassifying responses owing to some degree of subjectivity.

## BACKGROUND

News media is a leading source of health and scientific information for the public.[1,2]

Adolescents and young people frequently encounter and share news and information through digital media.[3,4] According to Eurostat, more than two-thirds of young people access online news media regularly.[4] More than half also deliberately search for health information online, indicating health-related topics to be important for youth, especially for those aged 15 and above.[5]

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.[6-8] 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.[9,10] Knowledge about scientific methods and scientific concepts is assumed as a necessity for appraising the reliability of health claims.[11,12] Health literacy initiatives at schools might help develop students’ skills in appraising claims, and some have suggested that these skills might empower students to make informed decisions about health and wellbeing 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]<sup>16</sup> School science might therefore be a key learning area for developing adolescents’ proficiency to critically appraise health claims in the media. Further, 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.[19-21,23,25-27] Moreover, they rely on substitute credibility indicators such as expertise (e.g., researchers, journalists) and authors' use of scientific statements and prompts (e.g., "evidence-based" or "scientifically proven") without any in-depth conceptual understanding.[19,22,28] 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 peoples' 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.[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?"[35, p.107]

Our study aimed to address this question, using data from a national science achievement test of Norwegian 10<sup>th</sup> grade students. To assess students' proficiency in identifying and appraising a health claim, we developed an item designed as a brief news report of a fictitious scientific study. As recognising a claim is a prerequisite for appraising the claim,[20,21,36] we asked the student to first identify the claim – the conclusion reported in the fictitious study. As requesting further evidence is a hallmark of critical appraisal when encountering claims,[37] we assessed whether and what type of information students requested about the reported study. Previous studies suggest that students, *if* they request information, usually

emphasize methodological aspects of the reported research, the findings as such, and theoretical explanations of the findings.[24,25,27,36,38,39]

In Norway, grade 10 is the final year of compulsory education and most students are 15-year-olds. According to the PISA studies of 15-year-olds, Norwegian students perform just above the OECD average in science.[40, p.44] Around 80% of Norwegian students perform at or above the level of scientific literacy associated with being proficient in identifying conclusions from simple data sets and hence identify scientific claims. Further, we can expect that just below 30% of Norwegian students possess skills necessary for appraising science-based claims.[16, p.285-6; 40, p.71] Building on knowledge from prior research and applying the national science achievement test of Norwegian 10<sup>th</sup> grade students as a measure of scientific literacy, we hypothesised that:

- (i) students who scored at the lower quartile on the scientific literacy measure are proficient in identifying a health claim among other competing textual information,
- (ii) students who scored 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

Cross-sectional, web-based science achievement test assessing a random sample of the 2013 cohort of 10<sup>th</sup> grade students in Norway.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Participants**

In 2013, the cohort of 10<sup>th</sup> grade students comprised about 64 000 individuals[41] distributed across 1238 schools.[42] Using random sampling, excluding special schools and international schools, two hundred 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 e-mail and telephone between 20<sup>th</sup> December 2012 and 6<sup>th</sup> February 2013. One class at each of ninety-seven schools – a total of 2229 students (50% females), completed the digitalized assessment during April-May 2013. We estimated the school/class average participation rate to 86%. Owing to technical shortcomings beyond our control, no data on students’ ethnicity was recorded. The Norwegian Directorate for Education and Training publishes the schools' final assessment grade in science. On a scale from 1-6, 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, science is a mandatory subject throughout compulsory education and is taught as an integrated subject called “natural science”. At the time of the survey, 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’.[43] The latter was introduced as a cross-cutting domain in the latest curriculum reform to ensure that *knowledge about science as a process* was integrated more systematically throughout science domains.

The national science achievement test measured students' proficiency in science based on the competence aims in the science curriculum for grade eight to ten with 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). In 2013, the main domains assessed were 'body and health' and 'diversity in nature'. The items were distributed across the science domains and cognitive domains as described in Supplementary file 1.

### **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 up to 4 points, were dichotomously scored selected response items. Accordingly, the science test data was analysed against the partial credit parameterization of the unidimensional Rasch model.[44,45] 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 complete 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 violations of unidimensionality or local independence was observed. The poorly discriminating item was discarded and the analysis was re-run.

The students completed the test within 80 minutes at school using a digitalized assessment tool.

The open-constructed 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 reduced 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.[37] 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 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), and to generate requests for information about the study 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. Responses to the news item allowed us to assess aspects of students’ *functional* and *critical* health literacy,[46] 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.[37] See Table 1 for an overview of the coding guide, including the taxonomy’s main scientific research categories (e.g. Methods). The coding guide was continually improved during field tests and clarified by including examples of authentic student responses (see Supplementary file 2 for a complete version of the guide).

**Table 1.** Overview of coding guide for the news item

Part 1: Health claim	
<b>Credited response</b>	<i>A response...</i>
Complete	providing a complete account of the claim (i.e. that a regular intake of corn reduces the risk of type II diabetes)
Mostly complete	providing a mostly complete account of the claim, with some significant words lacking (i.e. 'regular' and/or 'type II'), and/or refer to amount of intake (e.g. 'much' or 'more' corn).
<b>Non-credited response</b>	<i>A response...</i>
Wrong	where the claim relates to the topic of the news report, but is otherwise wrong
Vague	with no reference to <i>corn</i> and/or <i>type II diabetes</i>
Other	which is irrelevant or a 'don't know' response
Blank	
Part 2: Information requests	
<b>Credited response</b>	<i>A response relating to...</i>
Methods	<i>how</i> the study was conducted, including study design, subjects, procedures, and measurements
Data / Statistics	<i>what</i> was observed in the reported study, or about statistical tests used to analyse the data
Theory / Agent	<i>why</i> the reported effects might have occurred, including questions about the properties of the presumed causal agent and/or possible underlying mechanisms.
Social context	the credentials and bias related to who did the study or funded it and where it was conducted or published
Relevance	the importance or applicability of the study findings, or about the impact of the study
Related research	whether the findings have been replicated or fit results from previous research
Ambiguous	the study described in the news report that is ambiguous because it fits under two or more scientific categories
<b>Non-credited response</b>	<i>A response...</i>
Future studies	indicating the need for one or more <i>future</i> studies, either in general, or relating specifically to one of the scientific research categories (Methods, Data etc.)
Disbelief	indicating that the student doesn't believe that the study has been conducted
Wrong	relating to the topic of the news report, but is otherwise wrong
Vague	only vaguely referring to the scientific categories (Methods, Data etc)
Other	which is irrelevant or is a 'don't know'-response
Blank	

*Note.* 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 eighth 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".

One rater coded all student responses and consistency was evaluated by using an additional rater who coded twenty-five percent of the responses. Inter-rater agreement (ØG 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 low initial agreement rate was mainly owing to interpretation of responses that concerned the need for future studies (see specifications in Table 1). Our final decision was not to credit such responses, as the test item explicitly asked students to relate their requests to the study presented in the news report (item’s stem).

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 makes 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 make an educated assessment of the 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.

**Table 2.** News item scoring guide based on the coding guide for part 1 and 2

Credit	Score categories	Type of response
No credit	0	Wrong or vague health claim, irrelevant or blank response
Partial credit	1	Acceptable account of the health claim
	2	Acceptable account of the health claim and requests relating to one scientific research category (e.g. Methods)
	3	Acceptable account of the health claim and requests relating to two unique scientific research categories (e.g. Methods and Data)
Full credit	4 <sup>1</sup>	Acceptable account of the health claim and requests relating to three or more unique scientific research categories (e.g. Methods, Data, and Theory/Agent)

<sup>1</sup>None of the students made more than three unique requests for scientific information. Accordingly, full credit on the item was set to 4 score points.

The software package RUMM2030 was used for Rasch-analysis.[47] Using Rasch-analysis, one might 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.[48] The news item had five score categories (Table 2), and four *ordered* thresholds reflecting the difficulties of score 1 to 4 on the item. In Figure 1, we have located the four thresholds on the left side of the scale. The scale is made up of observable behaviours – the specific achievements associated with each threshold of the news item described in Table 2. These *observed* achievements were governed by the students' proficiency in science (scientific literacy) – the underlying but *unobservable* latent trait. On the right side of Figure 1 we have located the person (student) proficiencies associated with the 10<sup>th</sup> percentile, the quartiles and the 90<sup>th</sup> percentile. The possibility of locating item thresholds (difficulties) and person's proficiencies on the same logit-scale, is a benefit of Rasch-analysis. We used the information in Figure 1 to test both our hypotheses.

Missing data was handled using pairwise maximum likelihood estimator for the item location estimates - a so-called full information method. During field-trials, items displaying “differential item functioning” (DIF) for central person factors were revised or discarded. DIF means that e.g. males and females or minority and majority students with the *same* proficiency estimate have different probabilities of responding correctly. Hence, items displaying DIF are biased as gender and/or cultural background significantly influences students' responses. An example is an item assessing how hormones influence the menstruation cycle. This item probably uniformly favours females at all proficiencies along the latent trait scale.

[Insert Figure 1 here]

RESULTS

Two thirds of the students identified the health claim and 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), which equals the mean difficulty of the test items in the national achievement test. Accordingly, the average scientific literate student was likely able to identify the health claim in a brief news story. Students who reached the lower quartile on the scientific literacy scale were *not* likely to identify the claim, their proficiency (-0.5 logits) was 0.5 logits below the difficulty estimate of score point 1. Hence, hypothesis 1 is weakened.

**Table 3.** Proportion of students who identified the health claim of the news item and the proportions who requested different types of information for appraising that claim

	N (of 2229)	%
<b>Part 1: Health claim</b>		
<b>Credited response</b>	1420	64
Complete	710	32
Mostly complete	710	32
<b>Non-credited response</b>	809	36
Wrong, vague or other	415	18
Blank	394	18
<b>Part 2: Information requests</b>		
<b>Credited response</b>	652 <sup>1,2</sup>	29 <sup>2</sup>
Methods	376	17
Data /Statistics	189	9
Theory / Agent	146	7
Social context	57	3
Relevance	12	<1
Related research	9	<1
Ambiguous	31	1
<b>Non-credited response</b>	1577	71
Future studies	365	16
Disbelief	79	4
Wrong, vague or other	618	28
Blank	515	23

<sup>1</sup>Comprise all students who made one or more information requests, including 50 students who were not credited on part 1 (health claim). <sup>2</sup>As 154 students made more than one request, the number of specific requests exceeds the number of credited responses.

Less than one third of the students 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) – making one request for information, is rather close to the proficiency associated with the upper quartile (1.4 logits). Therefore, hypothesis 2 is partly strengthened. Students located even at or below the 90<sup>th</sup> percentile (2 logits) are *not* likely to score more than 2 points on the news item, i.e. make more than one request for information.

### Characteristics of students' information requests

As shown in Table 3, 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 the requests about Data/Statistics were rudimentary. In comparison, all requests about Theory/Agent were specific, e.g. 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).

**Table 4.** Characteristics of the requests related to Methods, Data and Theory (excerpts from Table 3)

Type of request	Examples of students' answers	No. of requests	%
<b>Methods</b>		<b>471</b>	<b>100</b>
Rudimentary	How was the study conducted?	80	17
Design	Did they use a control group? How long did the study last?	60	13
Agent delivery (procedure)	How much corn is necessary to eat? How often?	100	21
Participants	How many were tested? Who participated in the study? What was their eating habits?	230	49
Measures	Did they measure the participants' blood sugar?	1	>1



Data/Statistics		191	100
Rudimentary	We need the results.	111	58
Absolute nature of data**	For how many did this work? For how many did this not work?	32	17
Comparative nature of data*	By how much is the risk reduced? How large was the effect?	43	23
Duration of effect	Does it work over time?	5	3
Theory/Agent		149	100
Identification	We need to know what corn is composed of (nutrients)	2	1
Agent mechanisms	What ingredient in corn prevents diabetes type2? Why does corn prevent diabetes type 2?	117	79
(Side) effects	Is there any side effects of eating corn?	23	15
Alternative agents	Do other corn products have this effect?	7	5

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

Seventy-one percent 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 (0.64 logits).

DISCUSSION

We assessed how 10<sup>th</sup> grade students’ levels of scientific literacy corresponded to their ability to identify and critically appraise health claims in the news media, two important aspects of health literacy. The findings weakened our first hypothesis, as only the average scientific literate 15-year-old student leaving compulsory school is able to identify a clearly stated health claim in a rather simple news report of science. Students performing at the upper quartile of the scientific literacy measure are barely proficient at appraising that claim, namely making a request for evidence needed to determine its reliability. Accordingly, our second hypothesis was partly strengthened.

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.[24,25,27,36,38,39] To our knowledge, this is the first study investigating students' critical appraisal of a science-based health claim in the context of "mass education" with a large student sample at lower secondary school level. While about half of the invited schools declined to participate, the student participation rate was at an acceptable 86 %. Our analyses indicate that the sample average grade and gender distribution both were equal to the population, thus indicating generalizability of our findings. Furthermore, we have addressed a call for research on the *utility* of scientific literacy for critical appraisal of health claims.[35] The Rasch analysis demonstrated that the achievement test comprised a valid and unidimensional measure of scientific literacy, allowing us to evaluate properly the levels of scientific literacy necessary to involve in identifying and appraising health claims in news reports of science. To make students respond shortly and "on task", they were encouraged to write only two brief sentences. There is no evidence that this constrained their opportunities to make many requests for evidence regarding the claim. There is a potential for misclassification bias in the coding of students' responses as this process involved some degree of subjectivity. To decrease risk of bias, inter-rater agreement was investigated for 25% of the responses and consensus was reached for most cases. After reaching consensus, all similar responses in the entire data material were recoded.

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).[20,21,36] 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 average proficient in science managed to

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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.[20,21,36] Another issue not explored in our study, was whether the students perceived the claim as plausible or not. The former may be the case as plausibility has been found to provoke more methods questions.[36,38] For instance, students' occasional requests about procedures indicate 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 (e.g. the American scientists' affiliation), perhaps suggesting that students regarded science and scientists as authoritative and accordingly the claim plausible. This has also been noted in previous studies of students across educational levels.[20,21,23,25-27]

Our findings are concerning as they illustrate students' functional and critical health literacy at the end of compulsory school. Almost three quarters of the 10<sup>th</sup> grade students were unable to identify *and* critically appraise a health claim in a brief news report. Even the highest performing students mostly requested only one scientific criterion of an optimum of six broad criteria. Despite curricular mandates to develop scientific literacy and critical appraisal skills important for health literacy, it is clear that actual skills are underdeveloped and not taught in a way that improve students' understanding and assessment of health claims. 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, were lost in the need to emphasise factual knowledge on health topics.[49] Importantly, teachers did not acknowledge the relevance of teaching critical appraisal, or they lacked appraisal skills themselves. Thus, there is clearly a prospect for cross sector collaboration between healthcare and education

professionals and researchers to work together to enable laypersons to think critically about 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.[50,51] However, a recent cluster-randomised controlled trial showed promising effects of a cross-disciplinary developed intervention aimed at teaching primary school children to appraise claims about treatment effects.[52]

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 – *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 is 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.[6,7]<sup>7</sup> 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 (e.g. absolute vs relative risk).[11] Existing evidence suggests that such knowledge is better gained through teacher-guided investigations that allow students to reflect on adequate and inadequate experimental strategies, rather than through student-led hands-on or virtual experiments.[53] Importantly, teachers need to make explicit the link to critical reading and evaluation of health claims in the new(s) media. Several students suggested the conduct of a future study rather than requesting information about the reported study, often involving

1  
2  
3 themselves in doing so (e.g. “we have to test a number of people”). This perhaps supports the  
4  
5 notion that teachers enforce experimentation and hands-on activities without linking relevant  
6  
7 learning outcomes to reading critically appraising science presented in out-of-school contexts,  
8  
9 including media reports.[54,55] Finally, students hardly requested related research supporting  
10  
11 or disproving the claim. Accordingly, teaching could sensitise students to the limitations of  
12  
13 single studies, introducing the idea of systematic reviews.  
14  
15

16  
17 **ACKNOWLEDGEMENTS**  
18

19  
20 The authors would like to thank Jorån Østerholt Dalane (MSc) who, together with KSP,  
21  
22 developed the original version of the news brief item. We would also like to thank Bjørn  
23  
24 Vidnes, Anders Isnes, and Kirsten Fiskum who, together with ØG and LVN, developed all the  
25  
26 other science achievement test items.  
27  
28

29  
30 **AUTHOR CONTRIBUTIONS**  
31

32  
33 LVN, KSP and ØG developed a revised version of the news brief item and the coding guide,  
34  
35 and coded the student responses. ØG administered the survey, managed the data handling,  
36  
37 conducted the Rasch analysis and constructed the achievement scale. LVN wrote the  
38  
39 preliminary draft of the paper with significant contributions from ØG. The paper was  
40  
41 critically reviewed by KSP, BE and SF for important intellectual content.  
42  
43

44  
45 **FUNDING**  
46

47  
48 This research received no specific grant from any funding agency in the public, commercial or  
49  
50 not-for-profit sectors.  
51

52  
53 **ETHICAL APPROVAL**  
54

55  
56 The data set used in this paper did not include information that identified individuals, so  
57  
58 ethical approval was not required under Norwegian regulations.  
59  
60

## COMPETING INTERESTS

None declared.

## DATA SHARING STATEMENT

No additional information is available. A data file with a sample of the coded student responses (in Norwegian) is available on request. Please contact Øystein Guttersrud ([oystein.guttersrud@naturfagsenteret.no](mailto:oystein.guttersrud@naturfagsenteret.no)).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

REFERENCES

1. European Commission. Scientific research in the media (Special Eurobarometer 282). Brussels: European Commission, 2007.

2. Funk C, Gottfried J, Mitchell A. Science news and information today. Washington DC: Pew Research Center, 2017.

3. Anderson M, Jiang J. Teens, social media and technology 2018. Washington DC: Pew Research Centre, 2018.

4. Eurostat. Being young in Europe today. Luxembourg: Publications Office of the European Union, 2017.

5. Park E, Kwon M. Health-related Internet use by children and adolescents: systematic review. *J Med Internet Res* 2018;20:e120. doi: 10.2196/jmir.7731

6. Haber N, Smith ER, Moscoe E, et al. Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): A systematic review. *PLoS One* 2018;13:e0196346. doi: 10.1371/journal.pone.0196346

7. Schwitzer G. How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. *PLoS Med* 2008;5:e95. doi: 10.1371/journal.pmed.0050095

8. Sumner P, Vivian-Griffiths S, Boivin J, et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ* 2014;349:g7015. doi: 10.1136/bmj.g7015

9. Matthews A, Herrett E, Gasparrini A, et al. Impact of statin related media coverage on use of statins: interrupted time series analysis with UK primary care data. *BMJ* 2016;353:i3283. doi: 10.1136/bmj.i3283

10. Nagler RH. Adverse outcomes associated with media exposure to contradictory nutrition messages. *J Health Commun* 2014;19(1):24-40. doi: 10.1080/10810730.2013.798384

11. Austvoll-Dahlgren A, Oxman AD, Chalmers I, et al. Key concepts that people need to understand to assess claims about treatment effects. *J Evid Based Med* 2015;8:112-25. doi: 10.1111/jebm.12160
12. Schwitzer G. A guide to reading health care news stories. *JAMA Intern Med* 2014;174:1183-6. doi: 10.1001/jamainternmed.2014.1359
13. Grace M, Bay JL. Developing a pedagogy to support science for health literacy. *Asia-Pacific Forum on Science Learning and Teaching* 2011; 12(2).  
[https://www.eduhk.hk/apfslt/v12\\_issue2/foreword/index.htm](https://www.eduhk.hk/apfslt/v12_issue2/foreword/index.htm).
14. Sharples JM, Oxman AD, Mahtani KR, et al. Critical thinking in healthcare and education. *BMJ* 2017;357:j2234. doi: 10.1136/bmj.j2234
15. Harrison J. Science education and health education: locating the connections. *Studies in Science Education* 2005;41:51-90. doi: <https://doi.org/10.1080/03057260508560214>
16. OECD. PISA 2015 Assessment and Analytical Framework. Paris: OECD Publishing, 2017.
17. Millar R, Osborne J, editors. Beyond 2000: science education for the future. London: King's College London, 1998.
18. National Research Council. A Framework for K-12 Science Education. Washington DC: The National Academies Press, 2012.
19. Kolsto SD. 'To trust or not to trust...': pupils' ways of judging information encountered in a socio-scientific issue. *International Journal of Science Education* 2001;23:877-901. doi: <https://doi.org/10.1080/09500690010016102>
20. Norris S, Phillips LM. Interpreting pragmatic meaning when reading popular reports of science. *Journal of Research in Science Teaching* 1994;31:947-67. doi: <https://doi.org/10.1002/tea.3660310909>

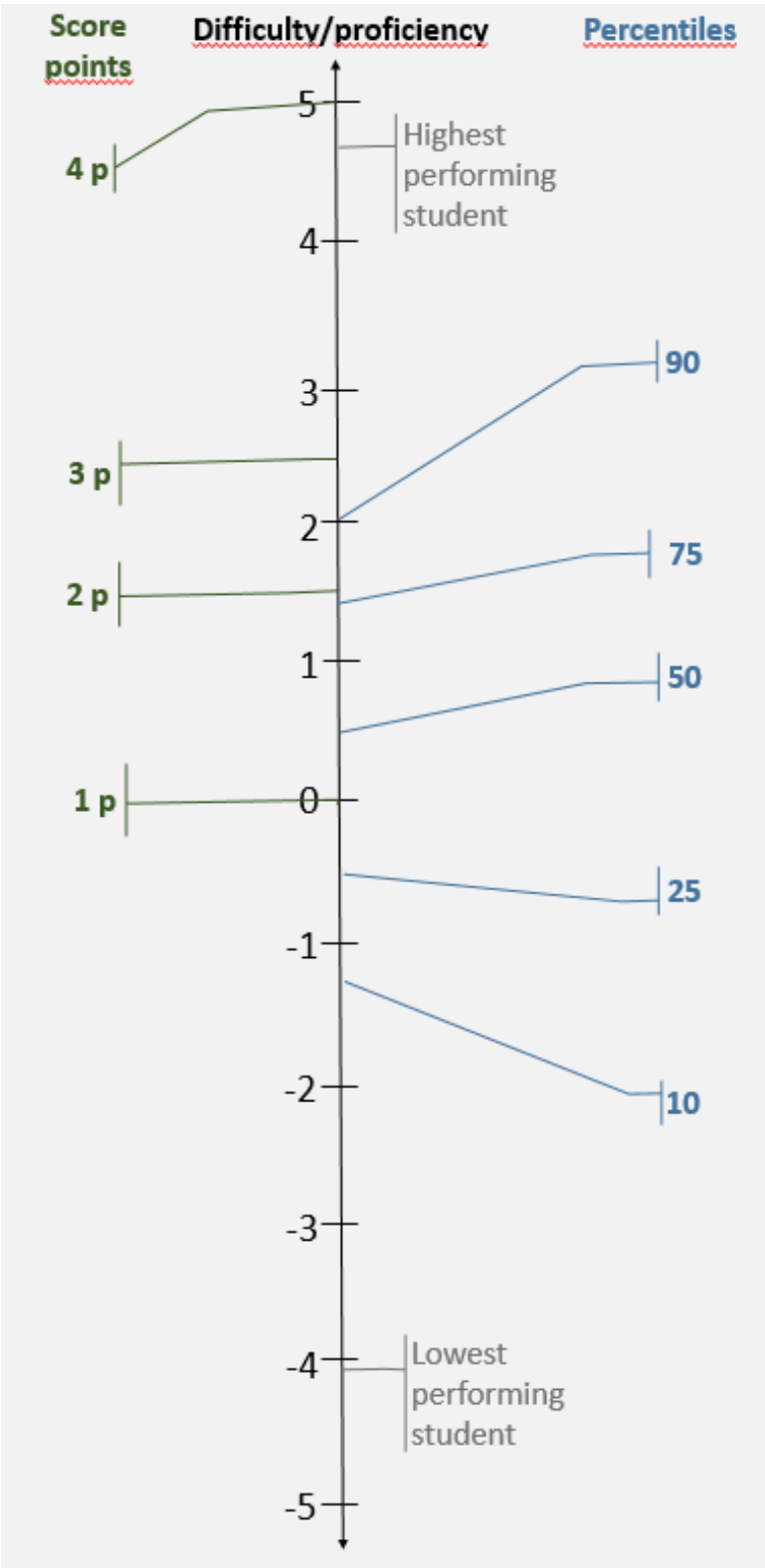


21. Norris SP, Phillips LM, Korpan CA. University students' interpretation of media reports of science and its relationship to background knowledge, interest, and reading difficulty. *Public Underst Sci* 2003;12:123-45. doi: <https://doi.org/10.1177/09636625030122001>
22. Oliveras B, Márquez C, Sanmartí N. The use of newspaper articles as a tool to develop critical thinking in science classes. *International Journal of Science Education* 2013;35:885-905. doi: 10.1080/09500693.2011.586736
23. Oliveras B, Márquez C, Sanmartí N. Students' attitudes to information in the press: critical reading of a newspaper article with scientific content. *Research in Science Education* 2014;44:603-26. doi: 10.1007/s11165-013-9397-3
24. Pettersen S. Critical thinking in Norwegian upper secondary biology education: the cases of complementary-alternative medicine and health claims in the media. *Nordic Studies in Science Education* 2005;1:61-71.
25. Pettersen S, Solberg J. Students of health sciences' evaluation of media reports of health research: a Norwegian study. In: Lewis J, Magro A, Simonneaux L, eds. Biology education for the real world: student - teacher - citizen; Proceedings of the IVth ERIDOB Conference; 2002 Oct 22-26. Toulouse-Auzeville: ENFA, 2003:293-306.
26. Phillips LM, Norris SP. Interpreting popular reports of science: what happens when the reader's world meets the world on paper? *International Journal of Science Education* 1999;21:317-27. doi: <https://doi.org/10.1080/095006999290723>
27. Ratcliffe M. Evaluation of abilities in interpreting media reports of scientific research. *International Journal of Science Education* 1999;21:1085-99. doi: <https://doi.org/10.1080/095006999290200>
28. Cusack L, Desha LN, Del Mar CB, et al. A qualitative study exploring high school students' understanding of, and attitudes towards, health information and claims. *Health Expect* 2017;20:1163-71. doi: 10.1111/hex.12562

29. Broder J, Okan O, Bauer U, et al. Health literacy in childhood and youth: a systematic review of definitions and models. *BMC Public Health* 2017;17:361. doi: 10.1186/s12889-017-4267-y
30. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12:80. doi: <https://doi.org/10.1186/1471-2458-12-80>
31. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. *Pediatrics* 2009;124:S265-74. doi: 10.1542/peds.2009-1162B
32. Fleary SA, Joseph P, Pappagianopoulos JE. Adolescent health literacy and health behaviors: A systematic review. *J Adolesc* 2018;62:116-27. doi: 10.1016/j.adolescence.2017.11.010
33. Guo S, Armstrong R, Waters E, et al. Quality of health literacy instruments used in children and adolescents: a systematic review. *BMJ Open* 2018;8:e020080. doi: 10.1136/bmjopen-2017-020080
34. Stellefson M, Hanik B, Chaney B, et al. eHealth literacy among college students: a systematic review with implications for eHealth education. *J Med Internet Res* 2011;13:e102. doi: 10.2196/jmir.1703
35. National Academies of Sciences E, and Medicine,. Science literacy: concepts, contexts, and consequences. Washington, DC: The National Academies Press, 2016.
36. Korpan CA. Science literacy: what do students know and what do they want to know? [Toronto]: Canadian Council on Learning, 2009.
37. Korpan CA, Bisanz GL, Dukewich TL, et al. Assessing scientific literacy: a taxonomy for classifying questions and knowledge about scientific research. Technical Report No. 94-1. Edmonton, Canada: University of Alberta, 1994.

38. Korpan CA, Bisanz GL, Bisanz J, et al. Assessing literacy in science: evaluation of scientific news briefs. *Science Education* 1997;81:515-32. doi: [https://doi.org/10.1002/\(SICI\)1098-237X\(199709\)81:5<515::AID-SCE2>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1098-237X(199709)81:5<515::AID-SCE2>3.0.CO;2-D)
39. Leung JSC, Wong ASL, Yung BHW. Evaluation of science in the media by non-science majors. *International Journal of Science Education, Part B* 2017;7:219-36. doi: 10.1080/21548455.2016.1206983
40. OECD. PISA 2015 Results (Volume I): Excellence and equity in education. Paris: OECD Publishing, 2016.
41. Norwegian Directorate for Education and Training. Grunnskolen informasjonssystem (The information system for primary and secondary schools) [program]. Oslo: NDfET, 2018.
42. Statistics Norway. Table 1. Primary and lower secondary schools. Pupils in primary and lower secondary school. Oslo: Statistics Norway, 2018  
<https://www.ssb.no/en/utdanning/statistikker/utgrs> (accessed 20 May 2018)
43. Norwegian Directorate for Education and Training. Natural science subject curriculum (NAT1-01). Oslo: NDfET, 2006.
44. Masters GN. A rasch model for partial credit scoring. *Psychometrika* 1982;47:149-74. doi: 10.1007/BF02296272
45. Rasch G. Probabilistic model for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research, 1960.
46. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int* 2000;15:259-67. doi: 10.1093/heapro/15.3.259
47. RUMM Laboratory Pty Ltd. RUMM2030 [program], 2011.
48. Andrich D. Rating formulation for ordered response categories. *Psychometrika* 1978;43:561-73. doi: <http://dx.doi.org/10.1007/BF02293814>

- 1  
2  
3 49. Nordheim L, Pettersen S, Flottorp S, et al. Critical appraisal of health claims: science  
4 teachers' perceptions and practice. *Health Educ* 2016;116:449-66. doi:  
5  
6 <http://dx.doi.org/10.1108/HE-04-2015-0016>  
7  
8  
9  
10 50. Cusack L, Del Mar CB, Chalmers I, et al. Educational interventions to improve people's  
11 understanding of key concepts in assessing the effects of health interventions: a systematic  
12 review. *Syst Rev* 2018;7:68. doi: 10.1186/s13643-018-0719-4  
13  
14  
15  
16 51. Nordheim LV, Gundersen MW, Espehaug B, et al. Effects of school-based educational  
17 interventions for enhancing adolescents abilities in critical appraisal of health claims: a  
18 systematic review. *PLoS One* 2016;11:e0161485. doi: 10.1371/journal.pone.0161485  
19  
20  
21  
22  
23 52. Nsangi A, Semakula D, Oxman AD, et al. Effects of the Informed Health Choices primary  
24 school intervention on the ability of children in Uganda to assess the reliability of claims  
25 about treatment effects: a cluster-randomised controlled trial. *Lancet* 2017;390:374-88. doi:  
26  
27  
28  
29  
30  
31  
32 10.1016/s0140-6736(17)31226-6  
33 53. Schwichow M, Croker S, Zimmerman C, et al. Teaching the control-of-variables strategy:  
34 A meta-analysis. *Developmental Review* 2016;39:37-63. doi:  
35  
36  
37  
38 <https://doi.org/10.1016/j.dr.2015.12.001>  
39  
40 54. Barton ML, Jordan D. Teaching Reading in Science: A Supplement to "Teaching Reading  
41 in the Content Areas Teacher's Manual (2nd Edition)". Aurora, Colorado: McREL, 2001.  
42  
43  
44 55. Ekborg M, Ottander C, Silfver E, et al. Teachers' experience of working with socio-  
45 scientific issues: a large scale and in depth study. *Research in Science Education*  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



# Supplementary file 1. Distribution of items across curriculum science domains and cognitive domains

Science subject domain	Number of items			
	All	Knowing	Applying	Reasoning
Diversity in nature	14	4 <sup>a</sup>	8	2
Body and health	18	7	4	7
The universe	4	1	1	2
Phenomena and substances	12	3	8	1
Technology and design	3			3
Budding researcher	4	1	1	2 <sup>b</sup>
<b>Total:</b>	<b>55</b>	<b>16</b>	<b>22</b>	<b>17</b>

<sup>a</sup>We deleted one item in this category owing to poor fit to the partial credit parameterization of the unidimensional Rasch model. <sup>b</sup>The news item was classified in this category.

Supplementary file 2. Coding guide with examples of student responses

Part 1: Health claim

Part 1 is coded using one variable in SPSS (v. 20). Use the codes as specified below.

CREDIT

ONE POINT

**Code 11:** Complete account of the claim: A **regular**<sup>1</sup> **intake**<sup>2</sup> of corn **reduces**<sup>3</sup> the **risk**<sup>4</sup> of type II diabetes.  
Accepted alternatives:  
<sup>1</sup>often, frequent(ly), daily, every day, several times a week, twice a week  
<sup>2</sup>eat (eating corn, people who eat corn).  
<sup>3</sup>decreases, lowers, diminishes, minimizes, prevents  
<sup>4</sup>chance, possibility, probability, likelihood, danger, (reduces) the number of people...

- Regular intake of corn reduces the risk of type II diabetes
- Eating corn reduces the risk of type II diabetes depending on how often you eat it.
- A daily intake of corn decreases the chance of developing type II diabetes.

**Code 12:** As code 11, but no reference to "regular" and/or "intake".

- Corn reduces the risk of type II diabetes
- Eating corn helps preventing type II diabetes

**Code 13:** As code 11, but no reference to "type II".

- A regular intake of corn reduces the risk of diabetes

**Code 14:** As code 11, but no reference to "regular" and "type II".

- Corn reduces the risk of developing diabetes

**Code 15:** As code 12 or 13, but reference to the **amount of intake** (more/much/specific amount of corn).

- People should eat more corn to avoid type II diabetes
- A certain amount of corn may prevent diabetes
- If you eat a lot of corn you decrease your chances for getting type II diabetes

NO CREDIT

**Code 01:** A response referring to the **wrong** claim, but the claim is related to the topic of the news report.

- A regular intake of corn may lead to type II diabetes (*misinterpretation*)
- By eating corn you don't get type II diabetes (*misinterpretation of "reduced risk"*)
- Researchers have found that eating corn cures type II diabetes (*intake of corn as treatment*)
- A regular intake of corn reduces the risk of cancer (*wrong diagnosis*)
- A dosage of corn may reduce your chances of developing diabetes (*eating corn as a «vaccine» for diabetes*)
- People eat too little corn
- The conclusion of the study is that there is little sugar in corn
- Diabetes type II is most common among humans
- That people with type II diabetes eat corn instead of exercising
- The conclusion is that the research findings are important

**Code 02:** A response referring to a **vague** claim.

- It means that the person eating corn have less risk (*no reference to type II diabetes*)
- That it reduces the risk of type II diabetes (*no reference to eating corn*)
- A regular intake reduces the risk
- Eat a lot of corn
- That you prevent diabetes

**Code 03:** Other responses.

- To find out how many that avoid type II diabetes by eating corn (*phrased as a problem statement, not a conclusion*)
- 1) Yes it does
- 1) Don't know.
- A conclusion in a study means that you don't necessarily find the answer doing a urine test (*trying to explain what a conclusion "is"*)

**Code 99:** Blank response to part 1.

## Part 2: Information requests

Part 2 is coded using eight variables in SPSS. Seven variables reflect the scientific research categories (Methods, Data, Theory/Agent etc.). Use code '1' or '0' to indicate whether the response include an acceptable request for information within the specific category. The eighth variable reflects different types of non-credited responses to Part 2, use the codes specified on page 3.

### CREDIT

Note: Credit on part 1 is a premise for credit(s) on part 2.

#### ONE POINT

Requests related to **one** of the following scientific research categories:

**Methods:** *How* the study was conducted, including research study design, subjects, procedures, and measurements.

- We need to know how the study was conducted (*rudimentary*)
- One should look at a control group and a main group, there is no information about this (*design*)
- How much corn should one eat? (*agent delivery*)
- How many have been tested? (*subjects*)

**Data/Statistics:** Presentation of research results or statistics - *what* was observed in the reported study, or about statistical tests used to analyse the data.

- We need the results (*rudimentary request*)
- Who reacted the most, women or men? (*comparative nature of the data for the dependent variable*)
- The results over time (*duration of effect*)
- Number tested and number that supported the hypothesis. (*absolute nature of the data for the dependent variable*)

**Theory/Agent:** Why the reported effects might have occurred, including questions about the properties of the presumed causal agent and/or possible underlying mechanisms.

- Why does corn reduce the risk of type II diabetes? (*mechanisms*)
- Is it dangerous to one's health to eat too much corn? (*agent effects*)
- Does this also apply to food products containing corn? (*alternative agents*)

**Social context:** The credentials and bias related to who did the research study or funded it and where it was conducted or published.

- Who researched this? (people)
- We need a reference to the study conducted by the researchers (*source of publication*)
- Do these results come from a reliable source? (*ambiguous: people OR source of publication*)

**Relevance:** The importance or applicability of the study findings, or about the impact of the study

- If intake [of corn] works in all people of all ages (*generalizability*)
- Is it possible for all people to get access to corn? (*practicality*)
- Approval from doctors/health authorities (*importance*)
- If people are willing to eat more corn to avoid diabetes (*utility*)

**Related research:** Whether the findings have been replicated or fit results from previous research, consensus/non-consensus among other researchers in the field.

- Have other researchers from other countries also arrived at this conclusion? (*supporting data*)
- How many have researched this? (*similar domain of study*)

**Ambiguous:** Requests that are relevant to the study described in the news report but that are ambiguous because they fit under two or more categories.

- How did they find out that corn reduces the risk of type II diabetes?
- How did they arrive at this conclusion?
- Where is the study conducted? [*could be Methods (research context) or Social context (Research institution)*]

#### TWO POINTS

Requests related to **two** scientific research categories, e.g.:

- How many participated in the study? How long did the study last? Result percentages? (*Methods; Data*)
- How was the study conducted? What ingredient in corn prevents type II diabetes? (*Methods; Theory*)
- How it was tested, how many was tested and for how long? Are the results coming from a reliable source? (*Methods; Social context*)

#### THREE POINTS

Requests related to **three or more** scientific categories., e.g.:

- We need the results and to see the research report. What [ingredient in] corn reduces the risk of diabetes? (*Data; Social context; Theory*)



NO CREDIT

- Code 01:** A response indicating the need for **future studies** that **specifically** relate to one or more of the **topic categories**
- We could ask people to eat corn to see if the number of people with type II diabetes decreases. We should find out how much corn they need to eat for this to work (*related to Methods*)
  - One needs clear results by testing humans (*related to Data and Methods*)
  - It needs thorough investigation over time with reliable reports that correspond to the findings (*related to Methods and Social context*)
  - More studies are needed to see if this works (*concerns Related research*)
  - It needs to be tested on groups to find out why this helps (*related to Theory/Agent*)
- Code 02:** A response indicating the need for **future studies in general**
- We need to do a study to test it
  - One should test it to see if the conclusion is valid
  - One can undertake a study to see that it reduces type II diabetes
- Code 03:** A response referring to a **vague** request for information
- How they figured it out
  - They need a form that shows it is actually true
  - Information about why it is like this
  - Proof / We need proof that it is true
  - A test
  - Approach
  - Research method
  - Sources of error
  - Examples / examples where it has helped
  - Documented effect
- Code 04:** A response indicating disbelief in the existence of the study
- We need information that the researchers have actually researched this
  - Has it been tested?
  - Results from a study (*not referring to 'the study'*)
- Code 05:** A **wrong or irrelevant** response, but the response is related to the topic of the news brief
- Since the Pan American Diabetes Research Group looks at this as an important research finding (*repeating text in news brief*)
  - Because it says that the study is conducted on a worldwide scale (*misinterpretation AND belief in the claim*)
  - The Pan American Diabetes Research Group is a credible name (*indicates belief in the claim*)
  - If the corn has been treated with a bactericidal liquid
  - Where corn is grown
- Code 06:** Other responses
- 2) Kebab
  - The risk of diabetes is reduced with a regular intake of corn. 2)
  - 1) Corn reduces the risk of type II diabetes. 2) Don't know
- Code 99:** Blank response to part 2.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study’s design with a commonly used term in the title or the abstract	1-2
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7-9
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants	7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	9-11
Bias	9	Describe any efforts to address potential sources of bias	11
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	9-12
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	11-12
		(b) Describe any methods used to examine subgroups and interactions	Not relevant
		(c) Explain how missing data were addressed	12
		(d) If applicable, describe analytical methods taking account of sampling strategy	Not relevant
		(e) Describe any sensitivity analyses	12
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	No information
		(c) Consider use of a flow diagram	Not relevant
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		(b) Indicate number of participants with missing data for each variable of interest	13
Outcome data	15*	Report numbers of outcome events or summary measures	13-15
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	13-15
		(b) Report category boundaries when continuous variables were categorized	Not relevant
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not relevant
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12
Discussion			
Key results	18	Summarise key results with reference to study objectives	15
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16-18
Generalisability	21	Discuss the generalisability (external validity) of the study results	16
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19

\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).

# BMJ Open

## 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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-028781.R1
Article Type:	Original research
Date Submitted by the Author:	06-Sep-2019
Complete List of Authors:	Nordheim, Lena Victoria; Western Norway University of Applied Sciences, Centre for Evidence-Based Practice Pettersen, Kjell Sverre; Oslo Metropolitan University, Faculty of Health Sciences Espehaug, Birgitte ; Western Norway University of Applied Sciences, Centre for Evidence-Based Practice Flottorp, Signe; Norwegian Institute of Public Health; University of Oslo, Institute of Health and Society Guttersrud, Øystein; University of Oslo, Faculty of Mathematics and Natural Sciences, Norwegian Centre for Science Education; Oslo Metropolitan University, Faculty of Health Sciences
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Evidence based practice, Paediatrics
Keywords:	Community child health < PAEDIATRICS, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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 V. Nordheim<sup>1\*</sup>, Kjell Sverre Pettersen<sup>2</sup>, Birgitte Espehaug<sup>1</sup>, Signe Flottorp<sup>3,4</sup>, Øystein Guttersrud<sup>2,5</sup>

<sup>1</sup>Centre for Evidence-Based Practice, Western Norway University of Applied Sciences, Bergen, Norway.

<sup>2</sup>Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway.

<sup>3</sup>Norwegian Institute of Public Health, Oslo, Norway.

<sup>4</sup>Institute of Health and Society, University of Oslo, Oslo, Norway

<sup>5</sup>Faculty of Mathematics and Natural Sciences, Norwegian Centre for Science Education, University of Oslo, Oslo, Norway

\*Corresponding author: Lena V. Nordheim, Centre for Evidence-Based Practice, Western Norway University of Applied Sciences, Bergen, Norway.

Inndalsveien 28

5020 Bergen

Norway

E-mail: [lvn@hvl.no](mailto:lvn@hvl.no)

Word count: 4 274

## 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 that 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.

**ARTICLE SUMMARY**

**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 to 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 digitalized 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 misclassifying responses owing to rater subjectivity.

## BACKGROUND

News media is a leading source of health and scientific information for the public,<sup>1, 2</sup> including adolescents and young people, who frequently encounter and share news and information through digital media.<sup>3, 4</sup> According to Eurostat, more than two-thirds of young people access online news media regularly.<sup>4</sup> More than half also deliberately search for health information online, indicating health-related topics to be important for youth, especially for those aged 15 and above.<sup>5</sup>

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.<sup>6-8</sup> 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.<sup>9, 10</sup> Knowledge about scientific methods and scientific concepts is assumed as a necessity for appraising the reliability of health claims.<sup>11, 12</sup> 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 wellbeing over the life course.<sup>13, 14</sup>

Some claim that a minimum level of scientific literacy is a prerequisite for developing health literacy.<sup>13, 15</sup> 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.<sup>16</sup> 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.<sup>16-18</sup>



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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.<sup>19-27</sup> Studies indicate that students tend to overestimate the certainty of scientific claims and accept them at face value.<sup>19-21, 23, 25-27</sup> Moreover, they rely on substitute credibility indicators such as expertise (e.g., researchers, journalists) and authors’ use of scientific statements and prompts (e.g., “evidence-based” or “scientifically proven”) without any in-depth conceptual understanding.<sup>19, 20, 28</sup> 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.<sup>29, 30</sup> Accordingly, these issues are hardly emphasised in measures and empirical studies of adolescents’ and young peoples’ health literacy.<sup>31-34</sup>

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.<sup>35 p. 107</sup> 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?”<sup>35 p. 107</sup> 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-year-olds. According to the PISA studies of 15-year-olds, Norwegian students perform slightly above the OECD average in science<sup>36 p. 44</sup> and approximately 80% and 30% perform at or above PISA proficiency level 2 and 4 in science, respectively.<sup>36 p. 320</sup> At level 2 students can *typically* “use common scientific knowledge to identify a valid conclusion from a simple data

set”<sup>36 p. 68</sup> and hence identify scientific claims – a prerequisite for appraising claims.<sup>20 21 37</sup> At level 4, students can *typically* “identify the evidence supporting a scientific claim” and draw on knowledge about scientific procedures (e.g., experimental designs) to justify conclusions.<sup>36 p. 72-4</sup> Hence, they can most likely request further evidence when encountering unsupported science-based claims, a hallmark of critical appraisal.<sup>38</sup> Previous studies suggest that students, *if* they request information, usually emphasize methodological aspects of the reported research, the findings as such, and theoretical explanations of the findings.<sup>24, 25, 27, 37, 39 40</sup> 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:

- (i) 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,
- (ii) 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<sup>41</sup> distributed across 1238 schools.<sup>42</sup> Using random sampling, excluding special schools and international schools, two hundred 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 e-mail and telephone between 20<sup>th</sup> December 2012 and 6<sup>th</sup> February 2013. One class at each of ninety-seven schools – a total of 2229 students (50% females), completed the digitalized 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’ socio-economic 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 e. On a scale from 1-6, 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’.<sup>43</sup> 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 eight to ten, 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 Supplementary file 1, the 2013 assessment emphasized 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 was analysed against the partial credit parameterization of the unidimensional Rasch model.<sup>44 45</sup> 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 minutes at school using a digitalized 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

reduce 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.<sup>38</sup> 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), i.e. that 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 - 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,<sup>46</sup> 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.<sup>38</sup> See Table 1 for an overview of the coding guide, including the taxonomy’s main scientific research categories (e.g. methods). We continually improved the coding guide during field tests and clarified it by including examples of authentic student responses (see Supplementary file 2 for a complete version of the guide).

**Table 1.** Overview of coding guide for the news item

Part 1: Health claim	
<b>Credited response</b>	<i>A response...</i>
Complete	providing a complete account of the claim (i.e. that a regular intake of corn reduces the risk of type II diabetes)
Mostly complete	providing a mostly complete account of the claim, with some significant words lacking (i.e. 'regular' and/or 'type II'), and/or refer to amount of intake (e.g. 'much' or 'more' corn).
<b>Non-credited response</b>	<i>A response...</i>
Wrong	where the claim relates to the topic of the news report, but is otherwise wrong
Vague	with no reference to <i>corn</i> and/or <i>type II diabetes</i>
Other	which is irrelevant or a 'don't know' response
Blank	
Part 2: Information requests	
<b>Credited response</b>	<i>A response relating to...</i>
Methods	<i>how</i> the study was conducted, including study design, subjects, procedures, and measurements
Data / Statistics	<i>what</i> was observed in the reported study, or about statistical tests used to analyse the data
Theory / Agent	<i>why</i> the reported effects might have occurred, including questions about the properties of the presumed causal agent and/or possible underlying mechanisms.
Social context	the credentials and bias related to who did the study or funded it and where it was conducted or published
Relevance	the importance or applicability of the study findings, or about the impact of the study
Related research	whether the findings have been replicated or fit results from previous research
Ambiguous	the study described in the news report that is ambiguous because it fits under two or more scientific categories
<b>Non-credited response</b>	<i>A response...</i>
Future studies	indicating the need for one or more <i>future</i> studies, either in general, or relating specifically to one of the scientific research categories (Methods, Data etc.)
Disbelief	indicating that the student doesn't believe that the study has been conducted
Wrong	relating to the topic of the news report, but is otherwise wrong
Vague	only vaguely referring to the scientific categories (Methods, Data etc)
Other	which is irrelevant or is a 'don't know'-response
Blank	

*Note:* 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".

One rater coded all student responses and consistency was evaluated by using an additional rater who coded twenty-five percent of the responses. Inter-rater agreement (ØG 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 was mainly owing to interpretation of responses that concerned the need for future studies (see specifications in Table 1). Our final decision was not to credit such responses, as the item’s stem explicitly asked students to relate their requests to the study presented in the news report.

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.

**Table 2.** News item scoring guide

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

*Note:* <sup>1</sup>No student made more than three unique requests for scientific information. Accordingly, full credit on the item was set to 4 score points.

The software package RUMM2030 was used for Rasch modelling.<sup>47</sup> Using unidimensional Rasch modelling, one may construct a scale and locate each item’s threshold(s) on that scale.

1  
2  
3 A dichotomously scored item has one threshold reflecting the difficulty of the item, and a  
4  
5 polytomously scored item has  $k-1$  thresholds reflecting the difficulty of its  $k$  score  
6  
7 categories.<sup>48</sup> The news item had five score categories (Table 2), and four ordered thresholds  
8  
9 reflecting the difficulties of each score. We located the four thresholds on the left side of the  
10  
11 scale in Figure 1. The scale was made up of observable behaviours – the specific  
12  
13 achievements associated with each threshold of the news item described in Table 2. These  
14  
15 observed achievements were governed by the students' proficiency in science (scientific  
16  
17 literacy) – the underlying but unobservable latent trait. On the right side of the scale (Figure  
18  
19 1) we located the person (student) proficiencies associated with the 10th percentile, the  
20  
21 quartiles and the 90th percentile. The possibility of locating item thresholds (difficulties) and  
22  
23 person's proficiencies on the same logit-scale, is a benefit of using Rasch modelling. We used  
24  
25 the information in Figure 1 to test both our hypotheses.

26  
27  
28 Missing data was handled using pairwise maximum likelihood estimation for the item  
29  
30 location estimates – a so-called full information method. During field-trials, items displaying  
31  
32 “differential item functioning” (DIF) for central person factors were revised or discarded. DIF  
33  
34 means that e.g. males and females or minority and majority students with the same  
35  
36 proficiency estimate have different probabilities of responding correctly. Hence, items  
37  
38 displaying DIF are biased as gender and/or cultural background significantly influences  
39  
40 students' responses. An example is an item assessing how hormones influence the  
41  
42 menstruation cycle. This item probably uniformly favours females at all proficiencies along  
43  
44 the latent trait scale.

45  
46  
47 [Insert Figure 1 here]  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



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), i.e. 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 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.

**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).

	N (of 2229)	%
<b>Part 1: Health claim</b>		
<b>Credited responses</b>	1420	64
Complete	710	32
Mostly complete	710	32
<b>Non-credited responses</b>	809	36
Wrong, vague or other	415	18
Blank	394	18
<b>Part 2: Information requests</b>		
<b>Credited responses</b>	652 <sup>1,2</sup>	29 <sup>2</sup>
Methods	376	17
Data /statistics	189	9
Theory / agent	146	7
Social context	57	3
Relevance	12	<1
Related research	9	<1
Ambiguous	31	1
<b>Non-credited responses</b>	1577	71
Future studies	365	16
Disbelief	79	4
Wrong, vague or other	618	28
Blank	515	23

*Note:* <sup>1</sup>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). <sup>2</sup>The total sum of requests will exceed 652 as 154 students made requests relating to more than one unique scientific category.

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 more than 2 points on the news item, i.e. 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. Based on analyses, we concluded that this technical deficit might have constrained the opportunities to make further requests for only 31 students.

### **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, e.g. 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).

**Table 4.** Characteristics of the requests related to methods, data and theory (specifications of Table 3)

Type of request	Examples of students' responses	No. of requests	%
<b>Methods</b>		<b>471</b>	<b>100</b>
Rudimentary	How was the study conducted?	80	17
Design	Did they use a control group? How long did the study last?	60	13
Agent delivery (procedure)	How much corn is necessary to eat? How often?	100	21
Participants	How many were tested? Who participated in the study? What was their eating habits?	230	49
Measures	Did they measure the participants' blood sugar?	1	>1
<b>Data/Statistics</b>		<b>191</b>	<b>100</b>
Rudimentary	We need the results.	111	58
Absolute nature of data*	For how many did this (not) work?	32	17
Comparative nature of data*	What is the reduction in risk? How large was the effect?	43	23
Duration of effect	Does it work over time?	5	3
<b>Theory/Agent</b>		<b>149</b>	<b>100</b>
Identification	We need to know what corn is composed of (nutrients)	2	1
Agent mechanisms	What ingredient in corn prevents diabetes type2? Why does corn prevent diabetes type 2?	117	79
(Side) effects	Is there any side effects of eating corn?	23	15
Alternative agents	Do other corn products have this effect?	7	5

Note: \*For the dependent variable (here: diabetes risk)

Seventy-one percent 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 literate 15-year-old student 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 generalizability of our findings. Although data on socio-economic status and ethnicity was unavailable for this study, previous studies have found these factors to predict science proficiency in Norwegian 15-year-olds.<sup>36, 49, 50</sup> For socio-economic status, however, the relationship with proficiency is relatively weak compared to 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.<sup>24, 25, 27, 37, 39, 40</sup> 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,<sup>51, 52</sup> we have addressed a call for research on the *utility* of scientific literacy for critical appraisal of health claims.<sup>35</sup> The analysis of responses to the news item provided useful information about *how* proficient

1  
2  
3 students were (i.e., the levels of scientific literacy necessary to involve in identifying and  
4 appraising health claims in news reports of science), and *what* kind of knowledge they  
5 possessed and applied when approaching such claims (i.e., which responses earned credits and  
6 which did not). Such in-depth knowledge of students’ thinking around a topic or task is an  
7 important outcome of secondary analyses of individual test items used in large-scale surveys,  
8 as it may formatively inform and develop teachers’ practices.<sup>53</sup>  
9

10  
11  
12 There were some limitations to this study. First, the test data were collected six years ago  
13 (2013), thus we acknowledge that a possible shift in students’ knowledge might have  
14 occurred. Nevertheless, our study is timely due to a major revision of the curriculum that will  
15 be implemented from autumn 2020. Second, to avoid response dependence between similar  
16 items, and accordingly violations of local independence in the data, the test comprised only  
17 one of the news item developed and field-tested. This prohibited us from evaluating whether  
18 variations in text dimensions (e.g., the claim’s plausibility and how familiar the students were  
19 with the health topic) could have influenced students’ information request – dimensions  
20 previously reported to impact on students’ critical engagement with news reports.<sup>37, 39</sup> For the  
21 same reason, and because the news item did not include any embedded attitudinal items, we  
22 were unable to assess whether important personal factors (e.g. interest in the health topic,  
23 belief in the claim, scientific attitude)<sup>24, 25, 37, 39, 40</sup> could have affected students’ requests.  
24  
25 Third, to make students respond shortly and “on task”, they were encouraged to write only  
26 two brief sentences. This might have constrained their opportunities to make several requests  
27 for evidence regarding the claim, although our analysis of incomplete responses due to the  
28 limit of 250 characters indicated this was probably not the case. Moreover, our findings  
29 resemble previous studies with regard to both the number and type of requests made.<sup>24, 25, 37, 39</sup>  
30  
31 Finally, while it is common practice to retain test items for re-use, this implies a lack of  
32 transparency in the test data used for this study.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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).<sup>20, 21, 37</sup> 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.<sup>20, 21, 37</sup> 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.<sup>37, 39</sup> 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 (e.g. 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.<sup>19-21, 23, 25-27</sup>

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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.<sup>54</sup> 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 think critically about health claims, as pointed out by Sharples et al.<sup>14</sup> Research is still scarce as to which interventions best improve students' ability to appraise health claims.<sup>55, 56</sup> 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.<sup>57</sup>

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.<sup>11</sup> 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.<sup>7</sup> 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.<sup>6, 7</sup> 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 (e.g.

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 inadequate experimental strategies, rather than through student-led hands-on or virtual experiments.<sup>58</sup> Importantly, teachers need to make explicit the link between experiments, critical reading, and evaluation 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 (e.g. “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 critically appraising science presented in out-of-school contexts, including media reports.<sup>59, 60</sup> Finally, students hardly requested related research supporting or disproving the claim. Accordingly, teaching could sensitise students to the limitations of single studies, introducing the idea of systematic reviews.

## ACKNOWLEDGEMENTS

The authors would like to thank Jorån Østerholt Dalane (MSc) who, together with KSP, developed the original version of the news brief item. We 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.

## AUTHOR CONTRIBUTIONS

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 significant contributions from ØG. The paper was critically reviewed by KSP, BE and SF for important intellectual content.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**FUNDING**

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**ETHICAL APPROVAL**

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

**COMPETING INTERESTS**

None declared.

**DATA SHARING 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 Guttersrud ([oystein.guttersrud@naturfagsenteret.no](mailto:oystein.guttersrud@naturfagsenteret.no)).

## REFERENCES

1. European Commission. Scientific research in the media (Special Eurobarometer 282). Brussels: European Commission, 2007.
2. Funk C, Gottfried J, Mitchell A. Science news and information today. Washington, DC: Pew Research Center, 2017.
3. Anderson M, Jiang J. Teens, social media and technology 2018. Washington, DC: Pew Research Centre, 2018.
4. Eurostat. Being young in Europe today. Luxembourg: Publications Office of the European Union, 2017.
5. Park E, Kwon M. Health-related Internet use by children and adolescents: systematic review. *J Med Internet Res* 2018;20:e120. doi: 10.2196/jmir.7731 [published Online First: 5 Apr 2018]
6. Haber N, Smith ER, Moscoe E, et al. Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): A systematic review. *PLoS One* 2018;13:e0196346. doi: 10.1371/journal.pone.0196346 [published Online First: 31 May 2018].
7. Schwitzer G. How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. *PLoS Med* 2008;5:e95. doi: 10.1371/journal.pmed.0050095 [published Online First: 30 May 2008].
8. Sumner P, Vivian-Griffiths S, Boivin J, et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ* 2014;349:g7015. doi: 10.1136/bmj.g7015 [published Online First: 17 Des 2014].
9. Matthews A, Herrett E, Gasparrini A, et al. Impact of statin related media coverage on use of statins: interrupted time series analysis with UK primary care data. *BMJ* 2016;353:i3283. doi: 10.1136/bmj.i3283 [published Online First: 30 Jun 2016].

10. Nagler RH. Adverse outcomes associated with media exposure to contradictory nutrition messages. *J Health Commun* 2014;19:24-40. doi: 10.1080/10810730.2013.798384 [published Online First: 15 Oct 2013].
11. Austvoll-Dahlgren A, Oxman AD, Chalmers I, et al. Key concepts that people need to understand to assess claims about treatment effects. *J Evid Based Med* 2015;8:112-25. doi: 10.1111/jebm.12160 [published Online First: 25 Jun 2015].
12. Schwitzer G. A guide to reading health care news stories. *JAMA Intern Med* 2014;174:1183-6. doi: 10.1001/jamainternmed.2014.1359 [published Online First: 7 May 2014].
13. Grace M, Bay JL. Developing a pedagogy to support science for health literacy. *Asia-Pacific Forum on Science Learning and Teaching* 2011; 12(2).  
[https://www.eduhk.hk/apfslt/v12\\_issue2/foreword/foreword7.htm](https://www.eduhk.hk/apfslt/v12_issue2/foreword/foreword7.htm) (accessed 4 Sept 2019).
14. Sharples JM, Oxman AD, Mahtani KR, et al. Critical thinking in healthcare and education. *BMJ* 2017;357:j2234. doi: 10.1136/bmj.j2234 [published Online First: 18 May 2017].
15. Harrison J. Science education and health education: locating the connections. *Studies in Science Education* 2005;41:51-90. doi: 10.1080/03057260508560214 [published Online First: 28 Mar 2008].
16. OECD. PISA 2015 Assessment and analytical framework. Paris: OECD Publishing, 2017.
17. Millar R, Osborne J, eds. Beyond 2000: science education for the future. London: King's College London, 1998.
18. National Research Council. A framework for K-12 science education. Washington, DC: The National Academies Press, 2012.

19. Kolsto SD. 'To trust or not to trust...': pupils' ways of judging information encountered in a socio-scientific issue. *International Journal of Science Education* 2001;23:877-901. doi: 10.1080/09500690010016102 [published Online First: 28 Mar 2008].
20. Norris S, Phillips LM. Interpreting pragmatic meaning when reading popular reports of science. *Journal of Research in Science Teaching* 1994;31(9):947-67. doi: 10.1002/tea.3660310909 [published Online First: 19 Aug 2006].
21. Norris SP, Phillips LM, Korpan CA. University students' interpretation of media reports of science and its relationship to background knowledge, interest, and reading difficulty. *Public Underst Sci* 2003;12:123-45. doi: 10.1177/09636625030122001 [published Online First: 1 Apr 2003].
22. Oliveras B, Márquez C, Sanmartí N. The use of newspaper articles as a tool to develop critical thinking in science classes. *International Journal of Science Education* 2013;35:885-905. doi: 10.1080/09500693.2011.586736 [published Online First: 26 Jul 2011].
23. Oliveras B, Márquez C, Sanmartí N. Students' attitudes to information in the press: critical reading of a newspaper article with scientific content. *Research in Science Education* 2014;44:603-26. doi: 10.1007/s11165-013-9397-3 [published Online First: 05 Feb 2014].
24. Pettersen S. Critical thinking in Norwegian upper secondary biology education: the cases of complementary-alternative medicine and health claims in the media. *Nordic Studies in Science Education* 2005;1:61-71. doi: 10.5617/nordina.485
25. Pettersen S, Solberg J. Students of health sciences' evaluation of media reports of health research: a Norwegian study. In: Lewis J, Magro A, Simonneaux L, eds. *Biology education for the real world: student - teacher - citizen*; Proceedings of the IVth ERIDOB Conference; 2002 Oct 22-26. Toulouse-Auzeville: ENFA, 2003:293-306.

26. Phillips LM, Norris SP. Interpreting popular reports of science: what happens when the reader's world meets the world on paper? *International Journal of Science Education* 1999;21:317-27. doi: 10.1080/095006999290723 [published Online First: 29 Jun 2010].
27. Ratcliffe M. Evaluation of abilities in interpreting media reports of scientific research. *International Journal of Science Education* 1999;21:1085- 99. doi: 10.1080/095006999290200 [published Online First: 20 Jul 2010].
28. Cusack L, Desha LN, Del Mar CB, et al. A qualitative study exploring high school students' understanding of, and attitudes towards, health information and claims. *Health Expect* 2017;20:1163-71. doi: 10.1111/hex.12562 [published Online First: 6 May 2017].
29. Broder J, Okan O, Bauer U, et al. Health literacy in childhood and youth: a systematic review of definitions and models. *BMC Public Health* 2017;17:361. doi: 10.1186/s12889-017-4267-y [published Online First: 27 Apr 2017].
30. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12:80. doi: 10.1186/1471-2458-12-80 [published Online First: 27 Jan 2012],
31. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. *Pediatrics* 2009;124 Suppl 3:S265-74. doi: 10.1542/peds.2009-1162B [published Online First: 5 Nov 2009].
32. Fleary SA, Joseph P, Pappagianopoulos JE. Adolescent health literacy and health behaviors: A systematic review. *J Adolesc* 2018;62:116-27. doi: 10.1016/j.adolescence.2017.11.010 [published Online First: 28 Nov 2017].
33. Guo S, Armstrong R, Waters E, et al. Quality of health literacy instruments used in children and adolescents: a systematic review. *BMJ Open* 2018;8(6):e020080. doi: 10.1136/bmjopen-2017-020080 [published Online First: 16 Jun 2018].

34. Stollefson M, Hanik B, Chaney B, et al. eHealth literacy among college students: a systematic review with implications for eHealth education. *J Med Internet Res* 2011;13:e102. doi: 10.2196/jmir.1703 [published Online First: 14 Dec 2011].
35. National Academies of Sciences, Engineering, and Medicine. Science literacy: concepts, contexts, and consequences. Washington, DC: The National Academies Press, 2016.
36. OECD. PISA 2015 Results (Volume I): Excellence and equity in education. Paris: OECD Publishing, 2016.
37. Korpan CA. Science literacy: what do students know and what do they want to know? [Toronto]: Canadian Council on Learning, 2009.
38. Korpan CA, Bisanz GL, Dukewich TL, et al. Assessing scientific literacy: a taxonomy for classifying questions and knowledge about scientific research. Technical Report No. 94-1. Edmonton, Canada: University of Alberta, 1994.
39. Korpan CA, Bisanz GL, Bisanz J, et al. Assessing literacy in science: evaluation of scientific news briefs. *Science Education* 1997;81:515-32. doi: [https://doi.org/10.1002/\(SICI\)1098-237X\(199709\)81:5<515::AID-SCE2>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1098-237X(199709)81:5<515::AID-SCE2>3.0.CO;2-D) [published Online First: 7 Dec 1998].
40. Leung JSC, Wong ASL, Yung BHW. Evaluation of science in the media by non-science majors. *International Journal of Science Education, Part B* 2017;7:219-36. doi: 10.1080/21548455.2016.1206983 [published Online First: 08 Jul 2016]
41. Grunnskolen informasjonssystem (The information system for primary and secondary schools) [program]. Oslo: The Directorate, 2018.
42. Statistics Norway. Table 1. Primary and lower secondary schools. Pupils in primary and lower secondary school. Oslo: Statistics Norway, 2018  
<https://www.ssb.no/en/utdanning/statistikker/utgrs> (accessed 4 Sept 2019)

43. Norwegian Directorate for Education and Training. Natural science subject curriculum (NAT1-01). Oslo: The Directorate, 2006.
44. Masters GN. A rasch model for partial credit scoring. *Psychometrika* 1982;47:149-74. doi: 10.1007/BF02296272
45. Rasch G. Probabilistic model for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research, 1960.
46. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int* 2000;15(3):259-67. doi: 10.1093/heapro/15.3.259 [published Online First: 01 Sept 2001].
47. RUMM Laboratory Pty Ltd. RUMM2030 [program], 2011.
48. Andrich D. Rating formulation for ordered response categories. *Psychometrika* 1978;43:561-73. doi: 10.1007/BF02293814
49. OECD. PISA 2006 Volume II: Data/Données. Paris: OECD Publishing, 2007.
50. Turmo A. Scientific literacy and socio-economic background among 15-year-olds—a Nordic perspective. *Scandinavian Journal of Educational Research* 2004;48(3):287-305. doi: 10.1080/00313830410001695745 [published Online First: 24 Jan 2007].
51. Ghaddar SF, Valerio MA, Garcia CM, et al. Adolescent health literacy: the importance of credible sources for online health information. *J Sch Health* 2012;82:28-36. doi: 10.1111/j.1746-1561.2011.00664.x [published Online First: 07 Dec 2011].
52. Khajouei R, Salehi F. Health literacy among Iranian high school students. *Am J Health Behav* 2017;41(2):215-22. doi: 10.5993/ajhb.41.2.13 [published Online First: 30 Apr 2017].
53. Angell C, Kjaernslie M, Lie S. Exploring students' responses on free-response science items in TIMSS. In: Shorrocks-Taylor D, Jenkins EW, Angell C, eds. Learning from

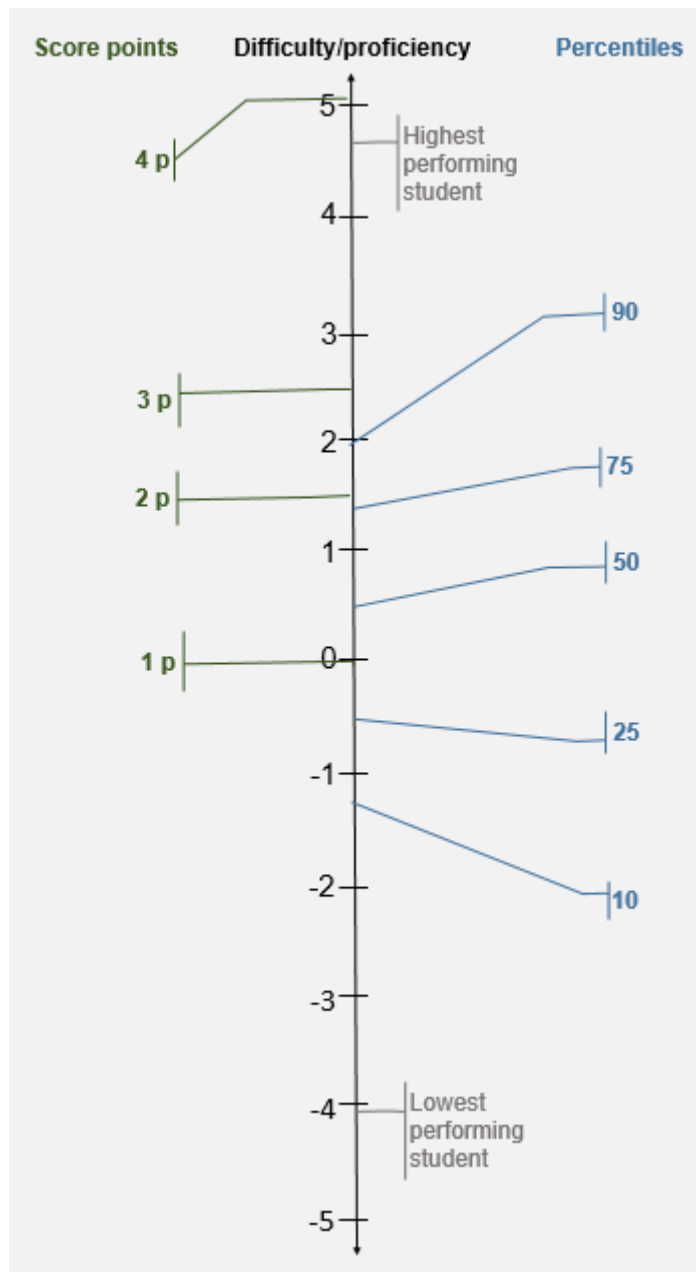
- others: international comparisons in education. Hingham, MA, USA: Kluwer Academic Publishers 2000.
54. Nordheim L, Pettersen S, Flottorp S, et al. Critical appraisal of health claims: science teachers' perceptions and practice. *Health Education* 2016;116:449-66. doi: 10.1108/HE-04-2015-0016 [published Online First: 01 Aug 2016].
55. Cusack L, Del Mar CB, Chalmers I, et al. Educational interventions to improve people's understanding of key concepts in assessing the effects of health interventions: a systematic review. *Syst Rev* 2018;7:68. doi: 10.1186/s13643-018-0719-4 [published Online First: 03 May 2018].
56. Nordheim LV, Gundersen MW, Espehaug B, et al. Effects of school-based educational interventions for enhancing adolescents abilities in critical appraisal of health claims: a systematic review. *PLoS One* 2016;11(8):e0161485. doi: 10.1371/journal.pone.0161485 [published Online First: 25 Aug 2016]
57. Nsangi A, Semakula D, Oxman AD, et al. Effects of the Informed Health Choices primary school intervention on the ability of children in Uganda to assess the reliability of claims about treatment effects: a cluster-randomised controlled trial. *Lancet* 2017;390(10092):374-88. doi: 10.1016/s0140-6736(17)31226-6 [published Online First: 26 May 2017].
58. Schwichow M, Croker S, Zimmerman C, et al. Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review* 2016;39:37-63. doi: 10.1016/j.dr.2015.12.001
59. Barton ML, Jordan D. Teaching Reading in Science: A Supplement to "Teaching Reading in the Content Areas Teacher's Manual (2nd Edition)". Aurora, Colorado: McREL, 2001.
60. Ekborg M, Ottander C, Silfver E, et al. Teachers' experience of working with socio-scientific issues: a large scale and in depth study. *Research in Science Education*



2013;43:599-617. doi: 10.1007/s11165-011-9279-5 [published Online First: 10 Feb 2012].

**FIGURE LEGENDS**

**Figure 1.** Difficulty of each score on the news item (left side) and the proficiencies associated with different percentiles (right side) on the scientific literacy scale



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Supplementary file 1. Distribution of items across curriculum science domains and cognitive domains

Science subject domain	Number of items			
	All	Knowing	Applying	Reasoning
Diversity in nature	14	4 <sup>a</sup>	8	2
Body and health	18	7	4	7
The universe	4	1	1	2
Phenomena and substances	12	3	8	1
Technology and design	3			3
Budding researcher	4	1	1	2 <sup>b</sup>
Total:	55	16	22	17

<sup>a</sup>We deleted one item in this category owing to poor fit to the partial credit parameterization of the unidimensional Rasch model. <sup>b</sup>The news item was classified in this category.

## Supplementary file 2. Coding guide with examples of student responses

**Part 1: Health claim**

Part 1 is coded using one variable in SPSS (v. 20). Use the codes as specified below.

**CREDIT****ONE POINT**

**Code 11:** Complete account of the claim: A **regular**<sup>1</sup> **intake**<sup>2</sup> of corn **reduces**<sup>3</sup> the **risk**<sup>4</sup> of type II diabetes.

Accepted alternatives:

<sup>1</sup>often, frequent(ly), daily, every day, several times a week, twice a week

<sup>2</sup>eat (eating corn, people who eat corn).

<sup>3</sup>decreases, lowers, diminishes, minimizes, prevents

<sup>4</sup>chance, possibility, probability, likelihood, danger, (reduces) the number of people...

- Regular intake of corn reduces the risk of type II diabetes
- Eating corn reduces the risk of type II diabetes depending on how often you eat it.
- A daily intake of corn decreases the chance of developing type II diabetes.

**Code 12:** As code 11, but no reference to "**regular**" and/or "**intake**".

- Corn reduces the risk of type II diabetes
- Eating corn helps preventing type II diabetes

**Code 13:** As code 11, but no reference to "**type II**".

- A regular intake of corn reduces the risk of diabetes

**Code 14:** As code 11, but no reference to "**regular**" and "**type II**".

- Corn reduces the risk of developing diabetes

**Code 15:** As code 12 or 13, but reference to the **amount of intake** (more/much/specific amount of corn).

- People should eat more corn to avoid type II diabetes
- A certain amount of corn may prevent diabetes
- If you eat a lot of corn you decrease your chances for getting type II diabetes

**NO CREDIT**

**Code 01:** A response referring to the **wrong** claim, but the claim is related to the topic of the news report.

- A regular intake of corn may lead to type II diabetes (*misinterpretation*)
- By eating corn you don't get type II diabetes (*misinterpretation of "reduced risk"*)
- Researchers have found that eating corn cures type II diabetes (*intake of corn as treatment*)
- A regular intake of corn reduces the risk of cancer (*wrong diagnosis*)
- A dosage of corn may reduce your chances of developing diabetes (*eating corn as a «vaccine» for diabetes*)
- People eat too little corn
- The conclusion of the study is that there is little sugar in corn
- Diabetes type II is most common among humans
- That people with type II diabetes eat corn instead of exercising
- The conclusion is that the research findings are important

**Code 02:** A response referring to a **vague** claim.

- It means that the person eating corn have less risk (*no reference to type II diabetes*)
- That it reduces the risk of type II diabetes (*no reference to eating corn*)
- A regular intake reduces the risk
- Eat a lot of corn
- That you prevent diabetes

**Code 03:** Other responses.

- To find out how many that avoid type II diabetes by eating corn (*phrased as a problem statement, not a conclusion*)
- 1) Yes it does
- 1) Don't know.
- A conclusion in a study means that you don't necessarily find the answer doing a urine test (*trying to explain what a conclusion "is"*)

**Code 99:** Blank response to part 1.

Part 2: Information requests

Part 2 is coded using eight variables in SPSS. Seven variables reflect the scientific research categories (Methods, Data, Theory/Agent etc.). Use code '1' or '0' to indicate whether the response include an acceptable request for information within the specific category. The eight variable reflects different types of non-credited responses to Part 2, use the codes specified on page 3.

CREDIT

Note: Credit on part 1 is a premise for credit(s) on part 2.

ONE POINT

Requests related to **one** of the following scientific research categories:

- Methods:** *How* the study was conducted, including research study design, subjects, procedures, and measurements.
- We need to know how the study was conducted (*rudimentary*)
  - One should look at a control group and a main group, there is no information about this (*design*)
  - How much corn should one eat? (*agent delivery*)
  - How many have been tested? (*subjects*)
- Data/Statistics:** Presentation of research results or statistics - *what* was observed in the reported study, or about statistical tests used to analyse the data.
- We need the results (*rudimentary request*)
  - Who reacted the most, women or men? (*comparative nature of the data for the dependent variable*)
  - The results over time (*duration of effect*)
  - Number tested and number that supported the hypothesis. (*absolute nature of the data for the dependent variable*)
- Theory/Agent:** Why the reported effects might have occurred, including questions about the properties of the presumed causal agent and/or possible underlying mechanisms.
- Why does corn reduce the risk of type II diabetes? (*mechanisms*)
  - Is it dangerous to one's health to eat too much corn? (*agent effects*)
  - Does this also apply to food products containing corn? (*alternative agents*)
- Social context:** The credentials and bias related to who did the research study or funded it and where it was conducted or published.
- Who researched this? (people)
  - We need a reference to the study conducted by the researchers (*source of publication*)
  - Do these results come from a reliable source? (*ambiguous: people OR source of publication*)
- Relevance:** The importance or applicability of the study findings, or about the impact of the study
- If intake [of corn] works in all people of all ages (*generalizability*)
  - Is it possible for all people to get access to corn? (*practicality*)
  - Approval from doctors/health authorities (*importance*)
  - If people are willing to eat more corn to avoid diabetes (*utility*)
- Related research:** Whether the findings have been replicated or fit results from previous research, consensus/non-consensus among other researchers in the field.
- Have other researchers from other countries also arrived at this conclusion? (*supporting data*)
  - How many have researched this? (*similar domain of study*)
- Ambiguous:** Requests that are relevant to the study described in the news report but that are ambiguous because they fit under two or more categories.
- How did they find out that corn reduces the risk of type II diabetes?
  - How did they arrive at this conclusion?
  - Where is the study conducted? [*could be Methods (research context) or Social context (Research institution)*]

TWO POINTS

Requests related to **two** scientific research categories, e.g.:

- How many participated in the study? How long did the study last? Result percentages? (*Methods; Data*)
- How was the study conducted? What ingredient in corn prevents type II diabetes? (*Methods; Theory*)
- How it was tested, how many was tested and for how long? Are the results coming from a reliable source? (*Methods; Social context*)

THREE POINTS

Requests related to **three or more** scientific categories., e.g.:

- We need the results and to see the research report. What [ingredient in] corn reduces the risk of diabetes? (*Data; Social context; Theory*)

## NO CREDIT

**Code 01:** A response indicating the need for **future studies** that **specifically** relate to one or more of the **topic categories**

- We could ask people to eat corn to see if the number of people with type II diabetes decreases. We should find out how much corn they need to eat for this to work (*related to Methods*)
- One need to test it on several people (*related to Methods*)
- It needs thorough investigation over time with reliable reports that correspond to the findings (*related to Methods and Social context*)
- It need to be tested on groups to find out why this helps (*related to Theory/Agent*)
- One needs clear results by testing humans (*related to Data and Methods*)
- More studies are needed to see if this works (*concerns Related research*)

**Code 02:** A response indicating the need for **future studies in general**

- We need to do a study to test it
- One can undertake a study to see that it reduces type II diabetes
- One should test it to see if the conclusion is valid

**Code 03:** A response referring to a **vague** request for information

- How they figured it out
- They need a form that shows it is actually true
- Information about why it is like this
- Proof / We need proof that it is true
- A test
- Approach
- Research method
- Sources of error
- Examples / examples where it has helped
- Documented effect

**Code 04:** A response indicating disbelief in the existence of the study

- We need information that the researchers have actually researched this
- Has it been tested?
- Results from a study (*not referring to 'the study'*)

**Code 05:** A **wrong or irrelevant** response, but the response is related to the topic of the news brief

- Since the Pan American Diabetes Research Group looks at this as an important research finding (*repeating text in news brief*)
- The Pan American Diabetes Research Group is a credible name (*indicates belief in the claim*)
- Because it says that the study is conducted on a worldwide scale (*misinterpretation AND belief in the claim*)
- If the corn has been treated with a bactericidal liquid
- Where corn is grown

**Code 06:** Other responses

- 2) Kebab
- 1) Corn reduces the risk of type II diabetes. 2) Don't know
- The risk of diabetes is reduced with a regular intake of corn. 2)

**Code 99:** Blank response to part 2.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study’s design with a commonly used term in the title or the abstract	1-2
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7-9
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants	7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	9-11
Bias	9	Describe any efforts to address potential sources of bias	11-12
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	9-12
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	11-12
		(b) Describe any methods used to examine subgroups and interactions	Not relevant
		(c) Explain how missing data were addressed	12
		(d) If applicable, describe analytical methods taking account of sampling strategy	Not relevant
		(e) Describe any sensitivity analyses	12
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	No information
		(c) Consider use of a flow diagram	Not relevant
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		(b) Indicate number of participants with missing data for each variable of interest	13
Outcome data	15*	Report numbers of outcome events or summary measures	13-15
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	13-15
		(b) Report category boundaries when continuous variables were categorized	Not relevant
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not relevant
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	16
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	3, 16-17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16-19
Generalisability	21	Discuss the generalisability (external validity) of the study results	16
<b>Other information</b>			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	21

\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).