HØGSKOLEN I BERGEN

BERGEN UNIVERSITY COLLEGE

Master's thesis

A systematic review on the effect of school-based educational interventions for enhancing adolescents' abilities in critical appraisal of health claims

Malene Wøhlk Gundersen

Master in Evidence-Based Practice for Health and Social Sciences
Centre for Evidence-based Practice
Faculty of Health and Social Sciences
Submission date: May 15th 2015

(Page intentionally left blank)

A systematic review on the effect of school-based educational interventions for enhancing adolescents' abilities in critical appraisal of health claims

En systematisk oversikt over effekten av skolebaserte undervisningsintervensjoner for å fremme ungdommers kompetanse i kritisk vurdering av helsepåstander

Master's student: Malene Wøhlk Gundersen

Main supervisor: Lena V. Nordheim

Co-supervisor: Signe Flottorp

Submission date: May 15th 2015

Words: 13619

(Page intentionally left blank)

Preface

I would like to be able to say that I always have been concerned with source criticism. But, truth be told, it is hard to say if this interest gave me the inspiration to become a librarian, or if it was the librarianship that made me critical to the wealth of erroneous information in existence. Whether innate or acquired along the way, it was not until I began my masters' degree that I truly acquired the taste for critical appraisal of scientific health-related articles. The realisation of being able to appraise and expose poor research, even without any formal training in health studies, is a lesson I eagerly use, both professionally and personally.

In the fast paced, information overloaded society we inhabit today, the ability to understand what to trust and what not to trust is crucial. The abundance of dubious information and seemingly trustworthy facts can easily lead the reader to the wrong conclusion. However in most cases we get hints. Hints revealing that the information may not be as trustworthy as the author would like us to think. Some of the hints are fairly easy to spot with a moderate amount of common sense, but for others a certain degree of knowledge about science is necessary. A kind of knowledge that has to be acquired thorough learning.

When selecting the topic of this master's thesis I therefore chose to examine how abilities in critical appraisal most effectively could be enhanced. This in itself is a quite large topic, so specifying further was necessary. Since knowledge about science must be learned, it was natural to focus on an educational setting. When this was furthermore seen in connection to a topic relevant to each and every one of us, namely health, it became apparent that an appropriate, interesting and very important question to answer was what effect school-based educational interventions could have on enhancing adolescents' abilities in critical appraisal of health claims. That the work would eventually lead to an actual publication of the systematic review, and not just be "another" master's thesis, also appealed to me immensely.

During the immersion in educational research, a field previously unknown to me, I had the pleasure of working with my outstanding supervisor, Lena Nordheim. I had high expectations but learned much more. You guided and encouraged me from beginning to end academically, and offered support when academic pursuits were suddenly put into perspective. Thank you.

I would also like to thank co-supervisor Signe Flottorp for solving conundrums underway, Birgitte Espehaug for statistical and methodological contributions, and my employer, the Norwegian Directorate of Health, as well as my colleagues, for supporting me.

Finally I would like to thank my wife, Hege. You are my constant. You know.

(Page intentionally left blank)

Abstract

This master's thesis consists of two parts. Firstly an introductory part elaborating on the method and decisions made underway in undertaking a systematic review and secondly an article manuscript. There will be little overlap between the two parts, since the introductory part supplements and discusses the article manuscript.

Background: The ability to appraise data and evidence scientifically, one of several competencies OECD stresses as a major goal for science education in schools, requires knowledge about science. Enhancement of this ability in adolescence might entail better understanding of how to identify reliable health claims. Until now no secondary research has been conducted on this topic.

Objective: To identify, appraise and synthesize studies examining the effectiveness of school-based educational interventions aiming to enhance adolescents' abilities in critically appraising health claims.

Methods: A systematic review including randomised and non-randomised controlled trials and interrupted time series. To identify these, an extensive literature search was conducted in health-related and educational databases. Two reviewers independently performed study selection, data extraction and risk of bias assessment with adapted versions of the Cochrane Collaborations tools and resources.

Results: Of over 17000 identified references five studies published in six reports was included: one randomised controlled trial, one cluster-randomised controlled trial and three non-randomised controlled trials. One study compared different teaching modalities, while four studies compared educational interventions to instruction as usual. Risk of bias was predominantly rated as high. The results showed small, but statistically significant effects in favour of the interventions.

Conclusion: The small number of included studies, their heterogeneity and low methodological quality inhibits any firm conclusions on the effects of school-based interventions for enhancing critical appraisal abilities in adolescents. The results indicate a beneficial effect, but must be interpreted with caution.

Keywords: Adolescent (MeSH), Health literacy (MeSH), Education (MeSH), Critical appraisal

Sammendrag

Denne mastergradsoppgaven består av to deler. Først en innledningsdel som utdyper metoden valgene som er tatt i utarbeidelsen av den systematiske oversikten, dernest et artikkelmanuskript til den systematiske oversikten. Det vil være lite overlapp mellom de to delene siden innledningsdelen er ment som et supplement til den systematiske oversikten.

Bakgrunn: Evnen til å vurdere data og forskning vitenskapelig, en av flere kompetanser OECD vektlegger som et mål for naturfagundervisningen i skolen, krever kunnskap om vitenskap. Forsterkning av denne evnen i ungdomsårene kan tenkes å medføre en bedre forståelse av hvordan pålitelige helsepåstander identifiseres. Frem til nå har det ikke forelagt oppsummert kunnskap innen dette emnet.

Hensikt og problemstilling: Å identifisere, vurdere og oppsummere studier som undersøker effekten av skolebaserte undervisningsintervensjoner for å fremme ungdommers kompetanse i kritisk vurdering av helsepåstander.

Metode: En systematisk oversikt som inkluderer randomiserte og ikke-randomiserte kontrollerte studier and avbrutte tidsserier. Et omfattende litteratursøk ble foretatt i helsefaglige og utdannings-relaterte databaser. To personer foretok uavhengig studieseleksjon, dataekstraksjon og kritisk vurdering ved hjelp av tilpassede versjoner av Cochrane-samarbeidets verktøy og ressurser.

Resultater: Av over 17000 identifiserte referanser ble fem studier publisert i seks publikasjoner inkludert. En randomisert kontrollert studie, en kluster-randomisert studie og tre ikke-randomiserte studier. En studie sammenlignet to undervisningsintervensjoner, mens de fire andre sammenlignet en undervisningsintervensjon med vanlig undervisning. Kvaliteten på studiene er overveiende lav. Resultatene viste små, men statistisk signifikante resultater i favør av intervensjonene.

Konklusjon: Det lave antallet inkluderte studier, deres heterogenitet og svake metodologiske kvalitet gjør det vanskelig å trekke sikre konklusjoner om effekten av skolebaserte undervisningsintervensjoner for å fremme ungdommers kompetanse i kritisk vurdering av helsepåstander. Resultatene indikerer en positiv effekt, men må tolkes med forsiktighet.

Nøkkelord: Ungdom (MeSH), Helseforståelse (MeSH), Utdanning (MeSH), Kritisk vurdering

Contents

Appendices to introductory part

Article manuscript

List of tables, figures, appendices and supplementary material

1. Introduction

This master's thesis is a systematic review on the effects of educational interventions for enhancing adolescents' abilities to critically appraise health claims. The thesis consists of two parts: an introductory part and the article manuscript reporting the systematic review. There may be a slight overlap between the parts, but mainly they will supplement each other. The article manuscript can be read without the introductory part, though that is not the case for the introductory part, seeing that essential information would be missed.

The first part, the introductory part, describes the development of the systematic review in greater detail than allowed for in the article manuscript and offers additional information. This part will focus especially on the method used to conduct the systematic review and a discussion of strengths and weaknesses of said method.

The second part, the article manuscript, is structured in accordance with both the IMRAD-format and the author guidelines for the journal in which publication is intended.

1.1 Background

In a recent Norwegian study 1624 students attending 7th grade in 105 different schools were posed the question: "Can you trust all the information available on the internet?" 84% answers the question correctly. In the same study 852 students attending 9th grade in 38 different schools were given two statements: "When I find information on the internet I will check to see if it corresponds with information found elsewhere." And "When I find information on the internet I am attentive to where it comes from." 85.5% and 78.8% respectively, answered that they agreed or partially agreed (Hatlevik et al., 2013).

High as they may be, the numbers do not comment on why the partially agreeing students are only just that, nor do they comment on the ability of the students to evaluate the trustworthiness of the information. Considering this in connection with health claims, claims are not necessarily based on scientific reliable facts even though they may appear that way. Many health claims are derived from poorly executed research, pseudoscientific facts or biased expert opinions (Cooper et al., 2012 ; Glenton, Paulsen & Oxman, 2005). This makes the ability to properly assess scientific content and the reliability thereof significant. Simply paying attention to the sender may not suffice, which leads to the question of on what level the sender is recognised. When 1178 children and adolescents from 9 to 13 years of age were asked which sources gave the most wrong information about health, 35% answered TV, 36%

answered friends, 5% answered school and only 6% answered the Internet (Brown, Teufel & Birch, 2007).

The role of the patient today is no longer characterised by the unconditional following of doctors' rules and recommendations. Shared decision making and patient participation is not only a way of describing the health system, it is encouraged by the authorities, who also recognises the extent to which patients actively seek out health-related information of varying quality (Ministry of Health and Care Services, 2012, p. 49). Furthermore, by Norwegian law the patient has a right to be informed of and receive necessary information regarding treatments to enable the patients' role as an active participant when choosing between different options (Norway. Patients' Rights Act, 1999).

Youth in Norway has the right to make decisions concerning their own health at the age of 16 (Norway. Patients' Rights Act, 1999). Considering both the law and the role of the patient, it is possible to reason that the school could be an ideal setting for teaching adolescents not only to ask where the information comes from, but also how it came to be. That is, for health information and claims, asking and understanding the scientific processes behind the fact, information or claim. Health education in schools would be an appropriate setting for teaching adolescents to learn about evaluating the health claims they encounter through the media or on the internet (Manganello, 2008), a statement supported by the Institute of Medicine referring to the educational system as a primary point of intervention to improve a broader type of literacy, health literacy (Nielsen-Bohlman et al., 2004, p. 142).

According to the competence aims after year 10 in the Natural Science subject curriculum of the Norwegian school system, knowledge and critical appraisal of information regarding body and health is essential for the preservation of one's own physical and mental health. This is one of several reasons the student should be able to identify natural science arguments in newspapers as well as other media and evaluate the content in a critical manner (Norwegian Directorate for Education and Training, 2013). In addition, the Ministry of Education and Research (2004, p. 48) stresses the importance of and need for the development of digital competence, i.e. the combination of basic ICT skills and the ability find, appraise and interpret the information, and that this should be facilitated by the educational system.

One way of broadly describing one aspect of scientific literacy is as the understanding of how scientific knowledge is created and being able to use these principles in everyday situations to establish if appropriate scientific processes have been used as basis for a given statement (Laugksch, 2000). It has been questioned if the understanding of scientific content,

i.e. the facts and knowledge presented, is not in itself sufficient to the act of appraising the quality of this content. To be able to evaluate the quality of the content, knowledge about science is necessary in addition to the practice of critical thinking (Pettersen, 2001 ; 2003).

Every third year the Organisation for Economic Co-operation and Development (OECD) measures skills and knowledge in reading, mathematics and science of 15-year-old students worldwide in the Programme for International Student Assessment (PISA). When last performed in 2012 a sample of 510.000 students took part, representing the 28 million 15-year-olds globally (OECD, 2015). According to PISA scientific literacy is defines as:

"…the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen. A scientifically literate person, therefore, is willing to engage in reasoned discourse about science and technology which requires the competencies to:

***1. Explain phenomena scientifically**: Recognise, offer and evaluate explanations for a range of natural and technological phenomena.*

***2. Evaluate and design scientific enquiry**: Describe and appraise scientific investigations and propose ways of addressing questions scientifically.*

***3. Interpret data and evidence scientifically**: Analyse and evaluate data, claims and arguments in a variety of representations and draw appropriate scientific conclusions."* (OECD, 2013, p. 7)

All these three scientific competences are important for appraisal. However the latter competency (3) is of particular interest to this review. In order to achieve this competency a student also need knowledge about science, that is the methods used by scientist to obtain valid and precise data, and an understanding of the ways in which claims in science are developed and justified (OECD, 2013).

Compared with the other participating countries in the PISA, Norway scored average in mathematics, above average in reading and below average in science. While the Norwegian student scored an average of 495 points on the test, the OECD average was 501. This lists Norway as 22nd among 34 OECD countries, with a relative position between 19th and 26th. This performance is not significantly different compared to the results from the previous assessments from 2009 and 2006 (OECD, 2012). Considering that the test focused on the students' ability to read, assess and use information presented in scientific texts, the results corresponds with the attention that has been devoted to reading skills in every school curricula (Kjærnsli & Olsen, 2013, p. 173).

Considering the suggested gains scientific literacy may lead to when in a health-related situation or setting and the potential for an improvement in the science education in

school, the next question may then be about how this is to be achieved in the best possible way.

1.2 Previous research

During scoping and planning this systematic review no other systematic reviews or review articles on the same topic were identified. Some systematic reviews partially overlapping with this review were identified. These are described in the article manuscript, but mainly they examined, for instance one of the teaching methods used in one of the included studies in this review, or the most effective intervention to enhance consumers' online health literacy.

In addition to the research mentioned above Horsley et al. (2011) investigated the effect of teaching critical appraisal skills to health professionals. After including three studies (n=272) they found statistically significant improvements in knowledge in two of two reporting studies and statistically significant improvements in critical appraisal skills in two of three reporting studies. When assessed with GRADE, the authors found that there was low to very low confidence in the estimates. Even though this systematic review targeted health personnel with former training in the health sciences, the results was similar to those mentioning previous research in the systematic review-part.

In another review also targeting health personnel by Nabulsi et al. (2007) the authors developed a conceptual framework consisting of key areas in the teaching and practicing of evidence-based healthcare allowing them to investigate both the effectiveness of training and also identify gaps in the research. Four of the outcomes included were in some way connected to critical appraisal. No effect estimates were given, but the authors described small to moderate improvements in knowledge in most cases.

Qualitative studies explore similar topics. Higgins, Begoray and Macdonald (2009) describe the implementation of a health education curriculum for grade 10 students in high schools. They studied how the program affected the students' ability to access, understand, evaluate and communicate health information. While not focusing directly on evaluating health information, the study reveals that 66% of the students would use the internet to corroborate the reliability of information. Further the students express satisfaction with the programs ability to enhance their health literacy.

1.3 What this systematic review contributes with

Since no kind of review articles on the same topic was identified, it was presumed that this systematic review would provide the first summary of the available evidence on the effect of educational interventions for enhancing critical appraisal abilities in adolescents.

It was hoped that the results of the systematic review would allow multiple conclusions to be drawn. One conclusion would be regarding the effects of the included interventions, which could be of use to science teachers planning new curricula or lessons. Another conclusion, based on the identified studies, would be recommending directions for future research to consider, and thereby being of use to educational researchers.

2. Objective and review question

The objective of this systematic review was to identify, appraise and synthesize studies examining the effectiveness of educational interventions in schools that aim to enhance adolescents' abilities in critically appraising health claims. The specific review question was:

What is the effect of school-based educational interventions for enhancing adolescents' abilities in critical appraisal of health claims?

3. Method

A systematic review "attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question" (Green & Higgins, 2011b). When conducted in a systematic and explicit manner the result is an exhaustive summary of the available evidence relevant to a research question. Key features of a systematic review are clearly defined objectives with pre-defined inclusion and exclusion criteria, a systematic literature search attempting to find all eligible studies, followed by a transparent and rigorous process to select and extract data from studies, to assess the validity of study findings and present and synthesise them in a systematic manner (Green & Higgins, 2011b).

I primarily used the recommendations of the Cochrane Collaboration (Higgins & Green, 2011) to inform the choices taken throughout the review process. In the following sections I will describe and discuss how these steps were performed for this specific systematic review.

3.1 Scoping search

The first step in conducting this systematic review was to perform a scoping search to examine if systematic reviews already existed and thereby avoid duplication of efforts. In a scoping search the research question is usually wide and the search itself is not as thorough as it would be in a systematic review (Armstrong et al., 2011).

The scoping search failed to identify any kind of reviews that specifically targeted educational interventions in schools to promote critical appraisal in health issues. The initiation of a new systematic review was therefore considered reasonable and the process of specifying inclusion and exclusion began.

3.2 Review protocol

During autumn 2013 I prepared a preliminary project plan that described the background, objective and method for the review (unpublished). During the first stages of the review process I gained insight both into the review topic and methodological issues that necessitated updates of the protocol. The final review protocol was published in PROSPERO earlier this year (Gundersen et al., 2015).

3.3 Eligibility criteria

As mentioned earlier one of the central aspects of a systematic review is a set of clearly defined criteria that describe characteristics of the studies eligible for inclusion in the review. Several frameworks supporting this procedure exist (Davies, 2011). Because this review aims to assess the effect of an intervention, I used PICOS, a set of components stating which participants, interventions, comparisons, outcomes and study designs that are of interest (O'Connor, Green & Higgins, 2011a). The eligibility criteria themselves are described in the article manuscript (See article manuscript). In the next sections I further elaborate on the rationale behind some of the specifications.

3.3.1 Participants

The participants of interest were adolescents, but since authors from different fields may handle the either the designation itself or the age group differently a clarification was necessary. The WHO defines adolescence to be from 10 to 19 years of age (World Health Organization, 2013). However, the inclusion criteria for the participants' age was set from 11

years of age up to and including 18 years of age, the reason being that the US National Health Education Standard defines 6th grade, the equivalent of 11 years of age, as the level for when students should be able to assess the validity of health information (Joint Committee on National Health Education Standards, 2007).

No exclusions were made on basis of gender, ethnicity, socioeconomic status, academic achievement or the like.

3.3.2 Interventions and comparisons

School-based, or educational, interventions are per definition complex and may include several components (Campbell et al., 2000). This complexity can pose challenges when framing the review question or applying the PICOS framework to the review question. A "typical" Cochrane review will investigate the effect of one specific intervention on a clearly defined patient group with regards to listed outcomes. For complex interventions this level of specificity is not always possible, or even desired, to achieve. Squires, Valentine and Grimshaw (2013) recommend applying a broader scope and include similar interventions rather than limiting the inclusion criteria to one intervention only, as this allows identification of generalizable features across interventions. Furthermore, interventions should be specified as clearly as possible, while at the same time keeping in mind that complex interventions are often poorly described in the primary studies.

In this review no particular interventions were preferred over others, the only requirement was administration of the educational intervention in a school setting. Ryder's (2001) framework of learning aims for knowledge about science in compulsory school was used to broadly describe the content of interventions. Importantly, interventions including any of the learning aims were eligible for the review. Similarly, the topic of health was defined broadly and studies using health only as context for teaching research methodology was included. Finally, the comparison group was not limited to interventions only within Ryder's framework, but included both instruction as usual and other interventions altogether.

A further note should be made with regard to Ryder's framework. The framework was chosen because it is empirically derived from a review of case studies of how lay people interact with science outside formal education (Ryder, 2001). Interestingly, but perhaps not surprising, a majority of these cases were health-related. When adjusting the framework to fit compulsory school science, Ryder (2002) removed learning aims relating to interpretation of data, including design characteristics of randomised controlled trials, because he considered

them conceptually too demanding for most students. Although recognising Ryder's point of view, understanding randomisation is important to critically review the evidence used to support claims about the effects of health interventions, a point made in the OECD PISA science framework (2013, p. 19). Accordingly, the category "interpretation of data" was added to the framework and studies that evaluated such teaching were included in the review (See Table 1).

Table 1 Ryder's framework of learning aims for knowledge about science in compulsory school science (cited from Ryder (2002 ; 2001))

Study design

Students should:

- Be aware of the range of methodologies used by scientists to collect data, e.g., *in vitro* and *in vivo* studies, blind and double-blind studies involving placebos, observational studies, and experimental studies involving control of variables.
- Understand that in experimental studies involving the control of variable, the choice of control variables impacts the validity of findings.
- Recognise that in population studies sample size and sampling bias have an impact on the validity of the findings.

Assessing the quality of data

Students should:

- Recognise that measurements carry an inherent variability and therefore do not provide unequivocal access to a 'true' value.
- Understand that an estimate of variability can be obtained from the spread found in repeated measurements.
- Recognise that if meaningful conclusions are to be drawn then communication of a measurement needs to be accompanied by an estimate of variability.

Interpretation of data[1]

Students should:

- Understand the terms correlation, causal link and causal mechanism[1].
- Be aware that when evidence for a correlation between two variables is presented a statistical estimate of the strength of the correlation should also be provided.
- Understand the distinction between proving a knowledge claim and using evidence to provide justification for a knowledge claim.
- Recognise that claims for a causal link can be justified using statistical data
- Recognise that randomised experimental studies, prospective studies and retrospective studies can provide justification (but not proof) for claims of a causal link.

Uncertainty in science

Students should:

- Appreciate that many scientific questions are not amenable to empirical investigation because of the number and complexity of variables which would need to be controlled in an experimental study, the long-time horizons involved. and/or restrictions on study design following from ethical considerations.
- Understand that since proof is often unattainable, decisions may need to be made on the basis of estimates of risk.

1. Science communication in the public domain

Students should:

- Understand the role of peer review in the publication of new findings.
- Be aware that the status, track record and funding source of scientists can influence how their interpretations of data are reported;
- Recognise that commercial organisations, scientists, government bodies and media reports often present measurements following from scientific investigation without any communication of the reliability or validity of these measurements.
- Appreciate that commercial organisations, scientists and government bodies can present unqualified reassurances which do not reflect the scientific uncertainties involved.
- Be aware that in describing disagreements between groups of scientists' media reports may provide limited consideration of the strength of each group's case.

[1]Not included in the 2002 original framework. [2]Learning aim included in the 2002 framework, under the Study design category

3.3.3 Outcomes

Bloom's taxonomy (Bloom et al., 1956) and Kirkpatrick's hierarchy (Kirkpatrick, 1967) are commonly used taxonomies for describing educational outcomes in formal and informal education. Both taxonomies include learning outcomes that can be classified as cognitive (knowledge and skills), affective (attitudes, satisfaction with training) and behavioural (use of knowledge and skills in everyday life or professional practice). Bloom's taxonomy was originally developed for formal (academic) education, and would thus be very relevant in a review about cognitive achievements. However, each domain is detailed into rather complex levels or sub-domains (Krathwohl, 2002). Kirkpatrick's levels are less fine-grained and were considered sufficiently detailed to describe the outcome domains for this review. Accordingly, critical appraisal abilities were categorised into the domains of knowledge and understanding, skills, behaviour, attitudes, and students' participation in the educational intervention. The main interest of this review was outcomes that could be classified within the cognitive and behavioural domain. Thus, studies evaluating affective outcomes (such as attitudes) were only included if also reporting on cognitive or behavioural outcomes.

Knowledge (e.g. understanding the principle of causality) and certain skills (e.g. identifying the cause and effect factor in a media report about a research finding) could be described as "surrogate" endpoints (O'Connor, Green & Higgins, 2011a) because they do not say anything about students' actual critical appraisal skill or behaviour. Nevertheless, they are prerequisite for actually performing critical appraisal (OECD, 2013 ; Ryder, 2001) and were considered relevant for the review.

3.3.4 Study designs

When we want to know if an an educational intervention or teaching method works or not, or which interventions or methods are most effective the randomised controlled trial (RCT) is the preferred study design (Torgerson & Torgerson, 2008, p. 1-3). Therefore, RCTs were an obvious choice for inclusion in the systematic review. In an RCT, participants are randomly allocated to receive an intervention (experimental group) or an alternative intervention or no intervention (comparison or control group). This is the is the only way to prevent systematic baseline dissimilarities between the groups and to be confident that any differences between them is due to the intervention, and not to confounding variables that might influence the outcome (O'Connor, Green & Higgins, 2011b ; Torgerson & Torgerson, 2008).

In the fields of social sciences and education RCTs are less used (The Campbell Collaboration, 2004). Inclusion of non-randomised controlled studies can be justified if the review question of interest cannot readily be answered by RCTs (Reeves et al., 2011). The initial scoping search indicated that RCTs on the review topic were possibly lacking. Thus, non-randomised controlled trials, defined as "an experimental study in which people are allocated to different interventions using methods that are not random" (Reeves et al., 2011), were included.

Interrupted time series, where a group of students was measured repeatedly before and after an intervention, were also included in the review. Observational studies, that is, studies where the researchers did not actively manipulate what happened to the groups, were excluded.

When studying interventions in schools, for instance the introduction of a new curriculum, it is often more feasible to allocate groups or clusters to the study arms rather than individual students (Torgerson & Torgerson, 2008 ; Higgins, Deeks & Altman, 2011b). For this reason, studies of clusters, including geographical areas, schools, classes, and teachers were eligible for the review.

A remark should be made with regard to the decision to include studies both with and without a pretest. Assessing of the outcome of interest in participants before they receive the intervention (pretest) is strongly recommended in both randomised and non-randomised controlled studies. This is by far the only way to evaluate if a change has really occurred at posttest (Shadish, Cook & Campbell, 2002, p. 136; 260). Notwithstanding this recommendation, studies without a pretest were still included because the knowledge, skills, and behaviour assessed in the review were considered to be rather advanced (Shadish, Cook & Campbell, 2002). However, the absence of a pretest was addressed as a separate domain when assessing risk of bias in studies (see section 3.7).

3.4 Literature searches

The quality of a systematic review is closely connected not only to the quality of the primary studies themselves, but also the quality of the literature search (Reed et al., 2005). If relevant studies remain unidentified this may bias the entire review, making conclusions invalid. Developing a literature search will always be a question about sensitivity, specificity and precision (Lefebvre, Manheimer & Glanville, 2011a). The topic for this review included concepts that were loosely defined and wide, such as "educational intervention" and "critical

appraisal". Accordingly, a more sensitive search approach was needed to identify as many relevant studies as possible, within the limits of available resources. For transparency and reproducibility, all search strategies are provided in the supplementary material for the article manuscript (See Supplementary material S1).

3.4.1 The development of the literature search in general

Including every aspect from PICOS in the search strategy is not necessarily desirable. In a search strategy for an intervention review it is generally sufficient to focus on participants (P), interventions (I) and study designs (S) (Lefebvre, Manheimer & Glanville, 2011b).

For optimal retrieval the search strategy should contain both subject terms from the databases' controlled vocabulary and text-words as used by the authors (Jenuwine & Floyd, 2004). Implementing this recommendation for the search section describing adolescents (P) resulted in an extremely high number of references, many of them irrelevant. After much testing, the final search strategy for the participants included subject terms and text-words describing ages, the equivalent grade levels and lower and upper secondary schools. This search section was similar throughout the databases, but adjustments had to be made to develop optimal searches in the subject specific databases.

The search section related to the educational intervention (I) was even more challenging to develop. I wanted to identify studies that described school-based educational interventions related to the teaching and learning of critical appraisal of health claims (I). Using MEDLINE as an example, the final search section consisted of subject terms and text words that described *education and teaching*, *sources of health information and claims*, and *critical appraisal*. It was necessary to use a variety of broad subject terms and text words. For instance, the concept of health literacy includes the ability to critically appraise health information (Sorensen et al., 2012). Thus, the search included a search filter for identifying health literacy studies developed by the National Library of Medicine (The Reference and Web Services Section, 2015). Moreover, *sources of health information* included obvious terms such as "consumer health information" but also broad terms for sources reporting research claims (e.g. the subject terms "Empirical research" and "Epidemiology").

In some of the databases I also included a search section describing relevant study types for the review (S), more specifically a *search filter*. A search filter is a predefined search strategy that aims to retrieve a particular set of records (Lefebvre, Manheimer & Glanville, 2011a). Methodological filters to limit the search to e.g. randomised controlled

trials or systematic reviews are commonly used. Filters can also be thematic such as the health literacy filter used in the intervention section of the search. I used a methodological filter originally created by the Effective Practice and Organisation of Care (EPOC) group in the Cochrane Collaboration.

I drafted a strategy for MEDLINE and ERIC together with my supervisor. In addition, an information specialist at The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI) peer reviewed the search strategy. She gave valuable comments to all thematic sections of the search and suggested additional terms for educational research to be incorporated in the methodological filter. An overview of the final search strategy MEDLINE is provided in Table 2. This strategy was adjusted to the other databases selected for the review.

Table 2 An overview of the MEDLINE search strategy

Sub-topic	**Search line**
PARTICIPANTS (P)	
Adolescents (ages)	1
Schools and grade levels	2-10
INTERVENTON (I): *EDUCATIONAL INTERVENTIONS* RELATED TO TEACHING/LEARNING *CRITICAL APPRAISAL*	
Education and teaching (ST[1], TW[2])	11-28
Sources of health information and claims (ST) AND Critical appraisal (ST)	29-54
Literacies relevant for critical appraisal combined[3] (e.g. scientific, health, information)	55-88
Sources of health information and claims (TW)	89-95
Literacies AND Sources of health information and claims (TW)	96
Critical appraisal of health information and claims (TW)	97-102
Critical appraisal of health information and claims – combined (ST+ TW)	103
P + I	
Adolescents (ages) OR Schools and grade levels AND Critical appraisal of health information and claims – combined (ST + TW)	104-106
STUDY DESIGNS (S)	
Methodological search filter	107 - 131
P + I + S	**132**

[1]SH: Subject terms. [2]TW: Textwords. [3]Including health literacy search filter - NLM

3.4.2 Bibliographic databases and adaptations of search strategies

No single database indexes articles in the field of health-related education. As a consequence reviewers that are concerned with both topics should consider searching both health-related databases and educational databases (Reed et al., 2005). No definite list of databases to include in a review exists, but Haig and Dozier (2003) suggest a core set of databases, most of them available for this systematic review. I search a combination of bibliographic databases related to health, education, information science and social science. All databases are listed in the article manuscript. For the health-related databases, the search strategy described in section 3.4.1 was the most precise strategy. Nevertheless, more precise results were retrieved in CINAHL using a modified, even more sensitive, strategy adapted for the educational databases. Due to poor indexing in the educational databases, the intervention section of the search strategy included terms that described health education, information, science and media, while a second part described critical thinking, literacy and evaluation.

The two social science and general science databases from Web of Science have no controlled vocabulary and therefore only allow text-word searching. I used the same text-words as in the other databases. The two databases from ProQuest were those thematically farthest from the review question. Thus, I developed a wider search strategy to avoid missing relevant studies, by focusing on school types and education (P) and literacy and critical thinking (I).

3.4.3 Literature searches performed in other sources

There is an association between having statistically significant and mainly positive results and getting research published. Hence completed studies with negative or less interesting results may never be published, a phenomenon referred to as publication bias (Lefebvre, Manheimer & Glanville, 2011a ; Egger & Smith, 1998). To minimise the risk of publication bias I therefore searched for grey literature and ongoing studies (Lefebvre, Manheimer & Glanville, 2011a), the former referring to less accessible research such as theses, conference papers, reports, and more. Sources for grey literature and ongoing studies are listed in the article manuscript.

Finally, reference lists of the studies included in the review were scanned for additional relevant studies, and citation searches were performed in ISI Web of Science, among other, by entering the reference of each included study into the index to examine if any studies published later on had cited it.

3.4.4 Managing references

I managed retrieved records using the bibliographic management system EndNote. After all searches were finalised I imported the records into an EndNote library and removed duplicates references. This was possible for all databases except a few resources used to identify ongoing studies and grey literature, which instead were copied into Word-documents for management and screening later in the review process.

3.5 Study selection

The process of selecting studies for inclusion in the review followed the literature searches. To ensure that no relevant studies are excluded wrongly, this process should ideally be performed independently by two review authors, using the pre-defined inclusion criteria (PICOS) (Higgins & Deeks, 2011). For pragmatic reasons the study selection for this systematic review deviated somewhat from this recommendation. Even after the duplicates were removed over 17000 references remained. To enable a more efficient workflow only I performed the initial screening alone.

If a record was clearly irrelevant, for instance by stating a different setting than schools, or an educational intervention that obviously did not meet the inclusion criteria, the study was excluded. Furthermore, if the record specified another population than students, adolescents, or 11 to 18 year olds the study was excluded. The sensitive literature search resulted in many obviously irrelevant references with titles such as: "Phase II study of ecteinascidin 743 in heavily pretreated patients with recurrent osteosarcoma". If any doubt, it was included to the second screening phase.

The second screening immediately followed the first one and was performed independently by myself and a second reviewer. We agreed beforehand that if consensus could not be reached, a third reviewer would be consulted. There were no disagreements at this stage, but several references were marked as unclear by one or both reviewers. We retrieved these in full text for further examination along with references that met the inclusion criteria.

Reading of full text was also performed independently by the same two reviewers and reasons for exclusion were noted for each excluded study. I contacted the study investigators in three instances where studies lacked sufficient information to determine eligibility (Higgins & Deeks, 2011). One study was excluded (Belland, Glazewski & Richardson, 2011) and one

(Leshowitz et al., 1993) was included based on the information provided by the authors. The third study was excluded as the author did not reply (Gegner, Mackay & Mayer, 2009).

A relatively large number of the references were obtained in full text because they either lacked an abstract, or the abstract did not provide sufficient details to decide on inclusion. The EPOC guidelines were followed when creating the table of excluded studies (Effective Practice and Organisation of Care (EPOC), 2015). Thus the table only comprises studies that readers in the field might have expected to see included; studies that were subject to great uncertainty among reviewers; studies where missing information had to be obtained from the study authors, and not every study assessed in full text.

3.6 Data extraction

Deciding upon what data to extract from studies lays the foundation for the review results. The data extraction form must be carefully developed to maintain a transparent method (Higgins & Deeks, 2011). I developed an initial form based on recommendations from the Cochrane Collaboration, the Best Evidence Medical Education (BEME) Collaboration, and a methodological article about systematic reviews of educational interventions (Higgins & Deeks, 2011 ; Hammick, Dornan & Steinert, 2010 ; Reed et al., 2005 ; Reeves et al., 2011 ; Cochrane Consumer and Communication Review Group, 2013).

One included study was used to pilot test the form. Pilot testing was done independently by me and a second reviewer. Data extractions were then compared and some elements in the form were adjusted. Although studies included outcomes other than those specified in the inclusion criteria, only data for outcomes relevant to the systematic review were extracted. A facsimile of the form is supplied in (See Appendix I).

Study authors should ideally be contacted if relevant study information is not present in the study report (Liberati et al., 2009). This proved necessary for all included studies and accordingly authors were contacted by e-mail. Authors supplied whatever information they could, but because of the time passed since the studies were carried out they were often not able to provide information or additional data. Details for one study are still awaiting (Hendricks, 2001 ; Hill, 1998).

3.7 Assessment of risk of bias in included studies

The next step of conducting a systematic review is to establish the internal validity of the included studies, more specifically whether any systematic biases exist in their results due to

the way the studies are conducted. A bias is defined as "a systematic error, or deviation from the truth, in results of inferences" (Higgins, Altman & Sterne, 2011). Notably, a study may have methodological flaws that do not affect the results negatively. Thus, we assess the risk of bias for each important outcome both within a study and across studies.

The risk of bias tool used for this particular review was developed in a process similar to the data extraction form, using the Cochrane risk of bias tool and a modified risk of bias form developed by the Cochrane Consumers and Communication Review Group (Higgins, Altman & Sterne, 2011 ; Ryan et al., 2013). Recently, Cochrane Collaboration published the ACROBAT-NRSI tool for non-randomised studies (Sterne et al., 2014). The tool is extensive and was not incorporated it in its entirety, but elements were extracted and integrated into the tool used in this review.

Because assessments will always be subjective, studies were judged independently by me and a second reviewer. We assessed each study on all domains in the tool and judged study outcomes as having low, unclear or high risk of bias. We also assessed the overall risk of bias for relevant outcomes across studies through GRADE evaluations (see section 3.8). In the following I give a brief summary of each domain in the risk of bias tool used for this review. A facsimile of the risk of bias tool is supplied in the appendices (See Appendix II).

3.7.1 Sequence generation and allocation concealment

Sequence generation refers to how participants are allocated to the intervention and comparison group(s). Random allocation is the only way to prevent confounding bias or allocation bias, which refers to forming comparison groups that are dissimilar in characteristics associated with the outcome of interest (Higgins, Altman & Sterne, 2011). Included studies that used a non-random allocation procedure, such as allocation by judgment of teachers or researchers, were always judged as having a high risk of bias for this domain.

Concealing the allocation procedure prevents foreknowledge of allocation of an individual by the researcher, participant or practitioner (e.g. teacher). Like adequate sequence generation, using appropriate procedures to hide the allocation to groups prevents confounding bias (Higgins, Altman & Sterne, 2011). In the present Cochrane risk of bias tool, bias due to inadequate sequence generation and allocation concealment is termed "selection bias". However, selection bias refers to bias in the selection of participants into the study and not biased allocation of recruited participants to groups (Sterne et al., 2014). We thus used "confounding bias" to refer to the two domains in the tool. All studies that were judged to

have high risk of bias for the sequence generation domain were equally judged to have high risk of bias for this domain.

3.7.2 Comparability in baseline characteristics and outcome measurements

If a baseline imbalance exists between the intervention and comparison groups for one or more variable linked an outcome, this may influence the effect estimates. Such an imbalance can happen by chance, but may also be a result of inadequate sequence generation or allocation concealment (Higgins, Altman & Sterne, 2011).

Baseline comparability comprises both characteristics and demographics such as age, gender, ethnicity and academic achievement as well as outcomes measured prior to the intervention to establish baseline values. Included studies were judged as having a high risk if obvious dissimilarities were present at baseline.

3.7.3 Blinding of students and education provides

Blinding is when measures are taken to ensure that one participant or a group of study participants are unaware of their group allocation in the intervention. The successful use of blinding minimises the chance of the participant knowing which intervention he or she received and thereby minimises bias. Knowledge of group allocation can affect the results by e.g. low expectations in the control group (Higgins, Altman & Sterne, 2011). Blinding can in general be difficult in educational research given the nature of the intervention and the often used study design of the non-randomised study (Sterne et al., 2014, p. 28). If blinding was not possible in the included studies, but would be assessed as having a low risk of bias if measures were taken to outweigh the lack of blinding.

3.7.4 Departures from intended interventions

This domain assessed if there were any differences between the groups in regards to departures from intended interventions. Such departures include cases where elements were administered in addition to the intervention and where elements in the intended intervention were not given. This assessment comprised co-interventions, i.e. interventions other than the studied intervention, contamination, i.e. if one study group receives an intervention intended for another study group, and fidelity of implementation, i.e. if the intervention was not implemented as intended (Sterne et al., 2014, p. 27). Included studies showing apparent

differences between groups or lacks in implementation would be judged as having a high risk of bias.

3.7.5 Blinding of outcome assessment

The purpose of blinding of the individuals performing outcome assessment is to prevent their possibility to introduce bias, which could be the case if they were aware of group allocation. The more subjective outcomes are especially vulnerable to bias if measures are not taken regarding this domain (Higgins, Altman & Sterne, 2011). Of special note in educational research is the importance of objective outcomes since the educator is often both developer and evaluator of a curriculum (Reed et al., 2005), which may result in lack of blinding of outcome assessor. Included studies would be judged as having a high risk of bias if blinding of outcome assessors was not done and this might have influence on the outcome measurement.

3.7.6 Incomplete outcome data and selective reporting

Incomplete outcome data occur when participants disappear during the study or are excluded from the analysis for a number of reasons and thereby raise the possibility of influencing the results. Of the reasons causing attrition participant withdrawal, missing attendance, insufficient answers and wrongful enrolment are just a few. It can be of special concern if the attrition rate is higher in one study group than the other, since this may affect the outcomes (Higgins, Altman & Sterne, 2011). Includes studies was judged as having a high risk of bias if the rate differed significantly between the groups of if outcomes were suspected to be affected.

Selective reporting happens when an author originally states what results are intended for publishing, but subsequently only reports on some of the reported outcomes or subsets. The concern is that results that are not statistically significant are chosen not to be published (Higgins, Altman & Sterne, 2011). If a protocol of the study has been published beforehand, comparing the stated outcomes with the outcomes reported in the final publication is a simple indication of whether selective reporting has occurred. The included studies were rated as having a low risk of bias if a confirming protocol was available or if author otherwise made it clear that no expected outcomes were left out.

3.7.7 Reliability and validity of outcome measures

As mentioned earlier Reed et al. (2005) stresses the need for objective outcome measures in educational research, especially in the cases where the author is also the researcher, developer and evaluator. The domains concerning the reliability and validity of outcome measures are of importance in educational interventions, since intervention-specific instruments are often necessary. When the use of already existing and documented instruments is not possible, new instruments should be assessed for their reliability, i.e. the degree to which they demonstrate consistency and reproducibility, and validity, i.e. the degree to which they measure what they are intended to measure (Reed et al., 2005). According to the Cochrane Handbook (Deeks, Higgins & Altman, 2011) it is important to know if the measurement scales used to measure cognitive abilities have been validated. They further state that researchers often adjust existing scales or instruments to suit their use better, but such changes necessitates an assessment of whether the instruments perform as intended.

3.7.8 Other bias and overall assessment

For the other bias domain there was no restriction as to what could be noted. Any factor that might have been the cause of bias and was not covered elsewhere in the tool was to be noted and assessed here. This could involve e.g. bias related to study design, fraudulence claims, inappropriate funding or any other problem.

The last domain in the risk of bias assessment was the judging of overall bias. Based on the eleven domains mentioned above, the overall bias was judged for each included study according to the Cochrane Handbook (Higgins, Altman & Sterne, 2011).

3.8 Data analysis

A meta-analysis is a statistical technique where the results from two or more studies are statistically combined. If the studies included in a systematic review allow it, the use of meta-analyses to statistically synthesize the results can contribute to both an increase in power and an improvement in precision. The use of forest plots when conducting the meta-analysis visualises the effect estimate from each study and additionally demonstrate what the combined effect estimate is. The sensibility in conducting meta-analyses depends on if the synthesized results could be misleading, which again depends on the heterogeneity between the studies (Deeks, Higgins & Altman, 2011).

The possibility of conducting meta-analyses was already described in the protocol. If the included studies were reasonably similar, the results would be combined in one or more meta-analyses as far as the results allowed. However, after the inclusion of studies ti the systematic review it was evident that the studies were quite heterogeneous with regard to participants, interventions and study designs. What's more, the data included was sparse and even when authors were contacted for additional data they were only to a limited extent able to provide this. Furthermore it was planned to analyse included randomised controlled trials separately, as recommended by the Campbell Collaboration (The Campbell Collaboration, 2004). Such a division will allow for the understanding of whether the results may have been influenced by methodological factors or whether the results act similarly across study designs.

Based on this the available data was summarised in tables and textually creating a narrative synthesis. Forest plots were created to visualise the effects for each outcome, since each forest plot contained only one outcome a fixed effects model was used. Mean differences was calculated for continuous outcomes and relative risk was calculated for dichotomous outcomes. For both types a confidence interval of 95% was used.

To grade the quality of the evidence GRADE (Grading of Recommendations Assessment, Development and Evaluation) was used to create Summary of Findings-tables in the Guideline Development Tool (McMaster University & Evidence Prime Inc., 2015). By using this approach it was possible to define the quality of the evidence not solely based on the risk of bias assessment, but also by displaying the confidence in the effect estimates in terms of directness, consistency and precision of the effect estimates (Schünemann et al., 2011). Even though GRADE was originally developed to assess the quality of the evidence and the express the strength of recommendations in health-related systematic reviews and clinical guidelines, it is just as applicable in the field of educational research.

Through an evidence profile consisting of different domains, the evidence for each outcome is assessed in terms of quality. Initially the study design is entered as either RCTs starting as high quality evidence, or observational studies starting as low quality evidence. Through the remaining domains the quality is maintained at original level, downgraded or in some cases upgraded. The following domains are risk of bias, consistency, directness, precision, and reporting (Schünemann et al., 2011).

Risk of bias corresponds with the usual risk of bias assessment. Through relevant domains each outcome is assessed for the chance of bias. In most cases when conducting a systematic review this process has been done earlier.

Consistency is meant to describe how homogenous the outcome is across the included studies. If the effect estimates are inconsistent in the studies, this may lead to downgrading of the quality, especially if the heterogeneity is unexplainable.

Directness is concerned with how similar the population, intervention, comparison and/or outcomes are compared to the eligibility criteria for the systematic review. If the results are not directly transferable, downgrading on the directness domain may be appropriate.

Precision deals with the imprecision of the results. Based on factors such as how large the study is and how much data is available, the imprecision can be assessed from the confidence intervals. When only one or a few small studies are available, it is common to see wide confidence intervals, indicating little precision in the effect estimates and leading to downgrading.

The last domain, reporting, gives the assessor the possibility to downgrade if publication bias is suspected.

4. Results

4.1 The included studies

The literature search and subsequent study selection process resulted in five studies published in six publications being included in the systematic review: one randomised controlled trial (Hendricks, 2001 ; Hill, 1998), one cluster-randomised trial (Kaelin et al., 2007), and three non-randomised trials (Leshowitz et al., 1993 ; Derry et al., 1998 ; Steckelberg et al., 2009).

All but one study (Steckelberg et al., 2009) were published in the US. Interventions across studies varied considerably in how they were delivered, their intensity and duration but all of them where addressing causality as a main topic (Hendricks, 2001 ; Hill, 1998 ; Leshowitz et al., 1993) or as one of several topics (Kaelin et al., 2007 ; Steckelberg et al., 2009). By using Ryder's framework for knowledge about science to classify intervention topics I found that most studies dealt with topics of within the areas of study designs, interpretation of data and science communication. Only the study by Steckelberg et al. included topics within all categories in Ryder's framework. This was by far the most advanced educational intervention covering many aspects of evidence-based medicine such as randomized controlled trials, estimates of risks and interpretation of diagnostic tests. An overview of intervention topics across studies is provided in Table 3 on the next page. The

studies are described in more detail in the article manuscript and summarised study characteristics are available in Table 1 and 2 in the article manuscript. In the next sections I will briefly summarise the risk of bias assessments across the studies, give a summary of the results and additional details regarding the overall quality of the documentation for outcomes.

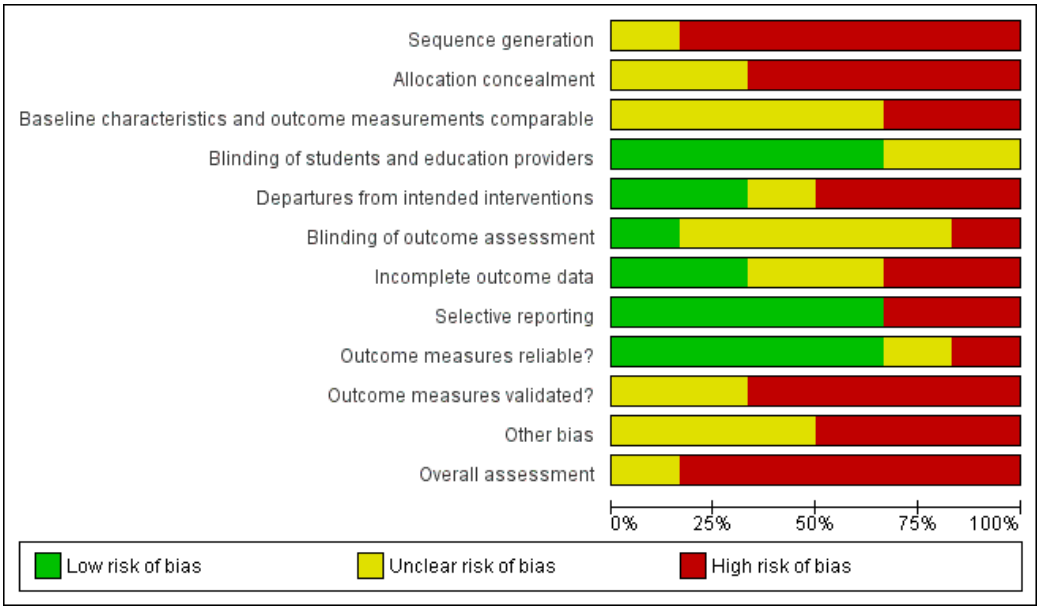
Table 3 Intervention topics across studies classified by Ryder's framework

	Study design	Assessing quality of data	Interpretation of data	Uncertainty in science	Science communication
Hendricks/Hill	Process of research; role of random assignment; control of variables	-	Causation vs correlation	-	Believability of claims in reports of science (real-world and constructed)
Derry 1998	Single-case observations versus randomised clinical trials, experimental control	-	Causation vs correlation	Uncertainty about effects of medical treatments; Probability (chance)	Funding issues (industry); governmental regulation of science
Kaelin 2007	Observational studies; confounding variables	-	Causation vs correlation	-	-
Leshowitz 1993	Control groups; confounding variables	-	Causation vs correlation	-	Believability of claims in reports of science (real-world and constructed)
Steckelberg 2009	Study designs evaluating effectiveness;	Confidence intervals	Critical appraisal of RCTs; correlation	Estimates of risks, precision/accuracy of diagnostic tests	Expert vs evidence-based information; misleading representation of health issues

4.2 Risk of bias in the included studies

The risk of bias in the included studies was generally high. As illustrated in the risk of bias graph the only domain with only low or unclear risk of bias was blinding of students and education providers. For the remaining domains sequence generation, allocation concealment and validation of outcome measures had a high prevalence of high risk of bias (See Figure 1).

Figure 1 Risk of bias graph



Several criteria for each of the included studies were judged as unclear. Even though authors were contacted for additional information and readily supplied the information when was available, the time passed since the studies were conducted made obtaining all the relevant information challenging in some cases and impossible in others.

Since the systematic review only generally describes the risk of bias assessment a more brief description of assessments made for each domain follows. In addition the risk of bias summary (See Figure 2 in article manuscript) and the table of risk of bias assessments made (See Supplementary material S5 in article manuscript) provide an overview of the assessments made.

Sequence generation was generally assessed as having a high risk of bias. For three of the studies the risk automatically became high since they did not use randomisation (Leshowitz et al., 1993 ; Derry et al., 1998 ; Steckelberg et al., 2009). One study was only able to include a small number of volunteer teachers and did therefore not achieve a true randomisation process (Kaelin et al., 2007). The last study examined two outcomes and due to the study design changing during the progress of the study, separate risk of bias assessments were made for each outcome. For the outcome related to causal reasoning randomisation was used, but sequence generation not mentioned, while for the outcome on transfer the original randomisation was broken, the risk of bias was thus assessed to be unclear and high respectively (Hendricks, 2001 ; Hill, 1998).

Allocation concealment was assessed as high for the three non-randomised studies (Leshowitz et al., 1993 ; Derry et al., 1998 ; Steckelberg et al., 2009). In one study the concealment method was not described (Kaelin et al., 2007), and in the last study the change in study design again entailed differentiation in the assessments. For the outcome related to causal reasoning not enough information was supplied to assess how allocation concealment had been done thus the risk of bias was assessed as unclear, and for the outcome on transfer the allocation was based on the researchers' decision resulting in a high risk of bias (Hendricks, 2001 ; Hill, 1998).

The *baseline characteristics and outcome measurements comparable* domain was assessed as having a high risk of bias for two studies since the groups were not comparable at baseline (Steckelberg et al., 2009 ; Leshowitz et al., 1993). The remaining three studies were assessed as having an unclear risk of bias because characteristics were given at population and not group level (Hendricks, 2001 ; Hill, 1998), because the characteristics were only described in text and pretest scores were not reported or described in regards to differences

between groups (Derry et al., 1998), and because no pretest scores were reported (Kaelin et al., 2007).

Blinding of students and education providers were assessed as having a low risk of bias in three studies since blinding was not possible and proper measures were taken in regards to this in the studies (Kaelin et al., 2007 ; Derry et al., 1998 ; Hendricks, 2001 ; Hill, 1998). In the two studies remaining blinding were not possible either, but in one the teachers were also the researchers (Leshowitz et al., 1993) and in the other low motivation in the intervention group could have been caused by the fact that the parallel classes had no more lessons before the summer holiday (Steckelberg et al., 2009). Risk of bias was assessed as unclear in both studies.

The *departures from intended interventions* domain was assessed as having a low risk of bias in one study based on careful documentation of the intervention (Hendricks, 2001 ; Hill, 1998). Another study did not provide sufficient information thus leading to the study being assessed as having unclear risk of bias (Leshowitz et al., 1993). The remaining three studies were all assessed as having high risk of bias. One study was lacking in intervention fidelity (Kaelin et al., 2007), one did not control sufficiently to exclude the possibility of contamination (Derry et al., 1998), and in one study the same teachers might have taught both intervention and control classes (Steckelberg et al., 2009).

Blinding of outcome assessment was assessed as having a low risk of bias in one study using Likert-scales and multiple-choice tests (Kaelin et al., 2007), three studies were assessed as having an unclear risk of bias because the interpretation of students' answers to the tests in each case required judgement (Derry et al., 1998 ; Hendricks, 2001 ; Hill, 1998 ; Steckelberg et al., 2009), and one study was assessed as having a high risk of bias because the interpretation of the students' answers required judgement and in addition the tests were scored by the teachers who also helped some students read the test (Leshowitz et al., 1993).

Incomplete outcome data was assessed as having a low risk of bias in one study based on the explanations given by the author (Hendricks, 2001 ; Hill, 1998), unclear in two studies since no information was given regarding absence in the control classes (Steckelberg et al., 2009 ; Leshowitz et al., 1993), and high in two studies on account of an administrative error (Derry et al., 1998) and on account of a substantial attrition unevenly distributed between groups (Kaelin et al., 2007).

Selective reporting was assessed as having low risk of bias in three studies since there was no reason to suspect selective reporting (Hendricks, 2001 ; Hill, 1998 ; Steckelberg et al., 2009 ; Leshowitz et al., 1993), in the remaining two studies risk of bias was assessed as

high since one study only provided subgroup results for subgroups (Kaelin et al., 2007) and one study reported only a scarcity of results (Derry et al., 1998).

For the *outcome measures reliable* domain three studies were assessed as having a low risk of bias since adequate measures were used and showed satisfying results (Hendricks, 2001 ; Hill, 1998 ; Kaelin et al., 2007 ; Leshowitz et al., 1993). One study was assessed as having an unclear risk of bias since no information on reliability measures was given (Derry et al., 1998). For the last study risk of bias was assessed as high since the instrument was under development and Rasch scalability was not yet achieved (Steckelberg et al., 2009).

For the *outcome measures validated* domain two studies were assessed as having an unclear risk of bias, one because validation was not mentioned (Leshowitz et al., 1993), and the other because one of two tests used in the study was conceptually more diverse from the instructional unit content than the other (Derry et al., 1998). The three remaining studies were assessed as having a high risk of bias either because only face validity was mentioned (Hendricks, 2001 ; Hill, 1998 ; Kaelin et al., 2007) or because the instrument was still under development (Steckelberg et al., 2009).

The *other bias* domain was open for any other type of bias that might have been introduced in the studies, but ended up almost exclusively describing lacking teacher characteristics and potential bias originating this aspect. Two studies were assessed as having an unclear risk of bias because either teacher behaviour was different between groups and the author did not describe reasoning behind teacher allocation (Hendricks, 2001 ; Hill, 1998) or because no information on teacher characteristics were given (Derry et al., 1998). The three remaining studies were assessed as having high risk of bias since teacher characteristics and demographics were not collected and factored in the analyses (Kaelin et al., 2007 ; Leshowitz et al., 1993 ; Steckelberg et al., 2009).

Based on these assessments one study was assessed as having an overall unclear risk of bias (Hendricks, 2001 ; Hill, 1998), while the remaining four studies all were judged as having a high risk of bias (Kaelin et al., 2007 ; Steckelberg et al., 2009 ; Derry et al., 1998 ; Leshowitz et al., 1993).

4.3 Effects of interventions

Each of the five studies administered different educational interventions developed by the researchers. The degree of similarity between the studies varied, making direct comparison of the different interventions challenging. Since all five studies consisted of different

combinations of e.g. pedagogical frameworks, teaching methods, content, number and duration of lessons, educators, and outcome measures, the results indicated the effect of each intervention, but did not permit assessment of which contributing components were the most successful.

The results presented in the included articles were in several cases very scarce. In some cases to the point that further data analysis was judged as not reasonable. Although authors were contacted to supply additional data, it was in many cases not collected originally or lost over the years since the research was conducted. This limitation naturally had consequences for the results presented in the systematic review.

The results from each of the five studies are only summarised below. A more detailed description is found in the article manuscript and the table of study characteristics (See article manuscript and Table 1 and 2 in article manuscript).

4.3.1 Educational interventions comparing different teaching modalities

Table 4 gives an overview of the study by Hendricks, the only study that compared two different teaching modalities (Hendricks, 2001 ; Hill, 1998). In this study a situated instruction model was compared to an abstracted instruction model for enhancing the students understanding of causality, subsequently transfer instruction was given to initiate spontaneous transfer.

Table 4 School-based educational intervention with results: Comparing different teaching modalities for enhancing adolescents' critical appraisal abilities

Study, *design*, allocation unit	**Students, setting**	**Intervention (n)**	**Comparison**	**Outcome [domain], *measure***	**Results**		**MD / RR (95%CI), p-value**
					Situated (SI)	**Abstracted (AI)**	
Hill 1998 / Hendricks 2001 (*Hendricks, 2001 ; Hill, 1998*) *Randomised controlled study with posttest only*; Students	220 students, 7th grade One lower secondary school, US	Situated instruction in causal reasoning (n=115)	Abstracted instruction in causal reasoning (n=105)	Understanding causality [Knowledge | Skills] *Selected and short-open response test: 0 to 15 points*	Mean posttest percentage score (12 p = 80%)		
					84.58 [SD: 17.88]	69.90 [SD: 19.76]	MD: 14.68 (9.86, 19.68) p<0.01
					Proportion mastering causality concept (Mastery = percentage score ≥ 80%)		
					88 of 115	47 of 105	RR 1.71 (1.35, 2.16) p<0.01
Transfer instruction: *Non-randomised controlled study with posttest only*; Class periods	194 of 220 participating students	SI + Transfer (n=60) SI + No Transfer (n=41)	AI + Transfer (n=34) AI + No Transfer (n=59)	Causal reasoning *Open-response test: 0 to 3 points*	Proportion mastering causal reasoning partially / completely (≥ 2 points)		
					SI + Transfer: 2 of 60	AI + Transfer: 0 of 34	-
					SI + No Transfer 0 of 41	AI + No Transfer 0 of 59	-

The mean difference between the groups after intervention was 14.68% (95% CI: 9.68 to 19.68, $p < 0.01$) in favour of the intervention. More students in the situated group mastered the concept of causality (percentage score ≥ 80) compared to the abstracted group (77 vs 45 students per 100, RR 1.71, (95% CI: 1.35 to 2.16, $p < 0.01$)), the risk ratio indicating that it is 1.71 times more likely that students who received situated instruction would master the concept of causality compared to the students who received abstracted instruction.

Where available data allowed further analysis, forest plots diagrams were created in Review Manager 5.3 (The Cochrane Collaboration, 2014). Using a fixed effects model the mean difference was calculated for continuous measures and risk ratio for the dichotomous measures. For both measures a 95% confidence interval was calculated (See Figure 2 and 3).

Figure 2 Forest plot illustrating mean difference for causal understanding

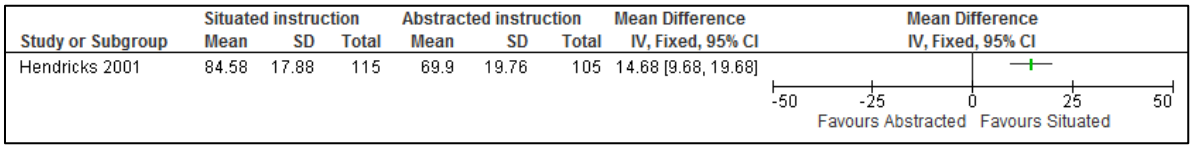
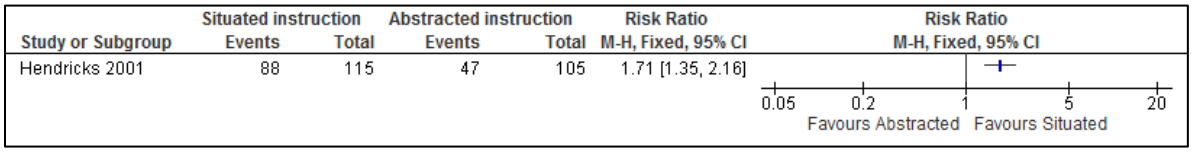


Figure 3 Forest plot illustrating risk ratio for mastering the concept of causality



4.3.2 Educational interventions

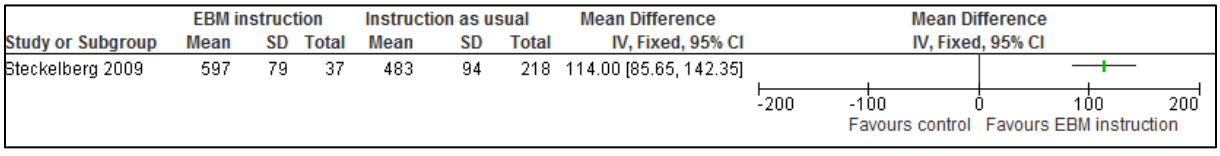
Table 5 on the next page gives an overview of the fours studies that compared educational interventions to instruction as usual. In all cases but one neither the data available in the articles nor through contact with the authors was sufficient for presenting results in a forest plot.

Table 5 School-based educational interventions compared to instruction as usual for enhancing adolescents' critical appraisal abilities, with results

Study, *design*, allocation unit [Ref]	Students, setting	Intervention topics and dosage	Comparison	Outcome, *measure*	Results		MD (95%CI). p-value
					Intervention	Usual instruction	
Derry 1998 *Non-randomised group study with pre- and posttest*; Teachers (Derry et al., 1998)	8th grade (no. of students not reported) 8 classes in one lower secondary school, US	Simulation gaming (role paly of legislation hearing) in causal reasoning (2 classes)	Instruction as usual (7 classes)	Causal reasoning *Open-response test: -1 to 13 points*	Mean adjusted posttest score: 3.61	2.27	Statistically significant difference (no p-values)
Kaelin 2007 *Cluster-randomised group study with pre- and posttest*; Teachers (Kaelin et al., 2007)	512 students in 7th grade 16 lower secondary schools, US	Epidemiology curriculum (n=378) *Subgroups: lessons (experience):* Intervention 1: 18 (1 experienced teacher) (n=88) Intervention 2: 16-18 (4 non-experienced) (n=197) Intervention 3: 6-10 (1 non-experienced) (n=93)	Instruction as usual (n=134)	Understanding epidemiology *Likert scale: 5 to 25 points (25 best)* *Multiple-choice test: 0 to 11 points*	Mean adjusted posttest score (reported for subgroups only): Int 1: 21.09[1] Int 2: 18.68 Int 3: 18.14 Int 1: 4.88[1] Int 2: 4.97[1] Int 3: 4.43	17.94 4.17	[1]Statistically significant difference (p<0.05)
Leshowitz 1993 *Non-randomised group study with posttest only*; Teachers (Leshowitz et al., 1993)	55 special (SE) and general education (GE) students in grade 7-12 5 classes in one lower and one upper secondary school, US	Causal reasoning instruction (n=22 SE students)	Instruction as usual (n=33 GE students)	Causal reasoning *Short-open response test: 0 to 6 points*	Unadjusted mean posttest score: 3.585	2.325	Statistically significant difference (p<0.01)
Steckelberg 2009 *Non-randomised group study with posttest only*; Classes (Steckelberg et al., 2009)	255 students in 11th grade 12 classes in upper secondary schools, Germany	Evidence-based medicine curriculum (n=37)	Instruction as usual (n=218)	Understanding EBM aspects *Multiple-choice and short-open responses*	Unadjusted mean person parameters (Rasch model): 597 [SD: 79]	483 [SD: 94]	114 (95%CI: 85.65, 142.35) (p<0.01)

Only the study by Steckelberg et al. (2009) provided sufficient details for presenting results in a forest plot diagram (See Figure 4). The forest plot illustrates a mean difference of 114 (95%CI: 85.65 to 142.35) in person parameter score between the intervention and control group.

Figure 4 Forest plot illustrating mean difference in person parameter score in understanding EBM aspects



4.3.3 Summary of findings

To assess the quality of the evidence and establish the degree of confidence in the effect estimates the GRADE-approach was used to create a summary of findings table (See Supplementary material). The difference in outcomes and study types prohibited summarisation in all cases but one. For each outcome the risk of bias assessment was incorporated and additional appraisal of directness, consistency, precision and risk of reporting bias was performed. Each outcome was downgraded by one or two points for study limitations, all but one outcome were downgraded one point for serious indirectness, only one outcome was downgraded for consistency and none were downgraded for risk of reporting bias (See Appendix III).

The quality of the evidence was for each outcome assessed as very low indicating a low confidence in the effect estimates.

5. Discussion

In the following sections a brief summary of systematic reviews in educational research will be given ahead of a discussion of which characteristics of both the included studies and the methodological limitations that may have affected the results of this systematic review.

5.1 Systematic reviews of educational research

In 1996 David Hargreaves, Professor of Education at University of Cambridge, opened his lecture by saying: "Teaching is not at present a research-based profession. I have no doubt that if it were, teaching would be more effective and more satisfying" (Hargreaves, 1996). He continued to say that educational research was non-cumulative because only few researchers attempt to create a body of knowledge that is tested, extended and replaced in a systematic way. Adding finally that education too needs evidence about what works with whom under which circumstances and with what effects, encouraging the educational researchers to look to medical research and evidence-based medicine.

In the early 2000s the establishing of the Campbell Collaboration and the broadened remit of the EPPI-Centre to undertake reviews in education displayed the movement towards establishing a closer connection between classrooms, research and policy (Bennett et al., 2005 ; The Evidence for Policy and Practice Information and Co-ordinating Centre, 2009). Nevertheless, while the augmentation was supported by the educational research community, the method of which it was done was debated. The advocates of systematic reviews

emphasised the characteristics of being objective, transparent, replicable and less vulnerable to bias, but the more sceptically inclined questioned the objectivity stating that even though the process might be transparent, the products would regardless be influenced by the values and judgments of the reviewers. Indeed not necessarily a limitation, but a premise needing to be recognised and utilised as to not induce a false confidence in the objectivity (Bennett et al., 2005 ; Eva, 2008). Also debated was the role of the "gold standard" RCT, the objection being that the limitations of the researchers' abilities to control all variables in educational contexts necessitates a more extensive range of methodological approaches to reach understandings or knowledge both in primary and secondary research (Bennett et al., 2005 ; Olson, 2004). In addition to this, ethical considerations have also been debated in relation to randomly assigning students to either receive or not receive an intervention (Reed et al., 2005).

This systematic review has strived to be transparent and replicable, and if not quite objective to a fault, then at least open and forthcoming about decisions made underway in the attempt to minimise bias. In the next two parts potential sources of bias originating from either the included studies or the methodological limitations of this review will be deliberated.

5.2 Challenges posed by the included studies

It is possible that both the quantity and quality of educational research has increased since 1996, but that is not to say that the diversity of interventions, participants and outcome measures has decreased. Each of the included studies, both the ones published before and after 1996, presented unique as well as similar characteristics, some of these are discussed below and some are discussed in the systematic review.

5.2.1 The teachers

According to Patrício and vaz Carneiro (2012) one of the challenges in educational research is to correctly identify the causal factors and the expected outcomes. The process of ascertaining an effect is not necessarily as easy as administering an intervention and examining the outcomes, but rather an intricate matter of differentiating the effect of the characteristics of the intervention from teacher effects. Such teacher effects can originate from communication skills, ability to motivate the students, use of humour or many other traits influencing perception and retaining. Hattie (2011) supports this association and not only stresses the importance of the interaction between the teacher and the student for learning, but argues that

the teachers' beliefs and commitments are the greatest influences on student achievement that we to some extent can control (Hattie, 2011, p. 22).

In one of the included studies the two teachers' administering the intervention was described as matched in general characteristics, but their instructional behaviour varied, as one of the teachers was described as more animated (Hendricks, 2001 ; Hill, 1998). In the remaining studies teacher characteristics was only sparsely described and the small amount of data available throughout the studies did not allow for any kind of analysis.

Considering the significance of the teachers, a more thorough description of both teacher characteristics and interaction with the students would have given a fuller picture of the administered intervention itself and made it possible to a greater extent to assess if the effects came from the intervention or other contributing factors.

5.2.2 The interventions

The heterogeneity that was expected when planning this systematic review was a contributing factor to the specification of interventions to those aimed at enhancing the critical appraisal abilities of claims and information about the human body and health. This could include different treatments, health conditions, physical and mental well-being or diseases. Furthermore, it was also decided not to include studies on regular health education interventions for outcomes such as smoking cessation or general subject matter knowledge, of which quite a few was encountered, e.g. in the field of genetics. A similar choice was made in regards to media-literacy interventions, which often involved some kind of health perspective. Such interventions would only be included when the critical examination of media messages contained aspects of knowledge about science.

Some of the excluded studies presented more of a challenge than others when eligibility was considered, as was the case with the study by Chowning et al. (2012). In this study the authors focused on teacher development and teaching materials to enhance the critical thinking skills of high school students through the use of bioethical case studies. Although the study was related both to health and critical thinking, the focus was of a more socio-scientific and ethical nature. To ensure that no studies were wrongfully excluded the dichotomy proposed by Kolstø (2006) was applied for socio-scientific issues. Kolstø distinguished between political questions concerning ethical, personal and social aspects of a scenario and risk questions concerning the appraisal of a disputed claim. For this systematic review studies dealing with risk questions were included, while studies dealing with political

questions were not. In cases such as the study by Chowning et al. the question was of a more political nature and therefore excluded (See Supplementary material 3).

Even though the outcomes in the includes studies, e.g. enhancing causal reasoning skills, knowledge in epidemiology or evidence-based medicine (EBM) aspects, were fairly related, the difference in approaches to intervention development resulted in considerable heterogeneity between the interventions and the teaching methods used. Whereas most studies utilised some kind of small group work (Derry et al., 1998 ; Hendricks, 2001 ; Hill, 1998 ; Steckelberg et al., 2009 ; Kaelin et al., 2007), only one study used class discussions (Leshowitz et al., 1993). And while some studies used extensive project work with mentoring (Derry et al., 1998) or without mentoring (Steckelberg et al., 2009), one study used investigations (Kaelin et al., 2007) and another used reflective activities (Hendricks, 2001 ; Hill, 1998).

This variety of teaching methods makes it close to impossible to identify which singular elements, if any, that succeeds in enhancing the skills and knowledge of the students. When compared to the work of John Hattie who synthesised over 900 meta-analyses to estimate the effect size of 150 influences on student achievement, the types of teaching methods integrated in the interventions of the included studies were spread wide apart. Where classroom discussion had a large estimated effect size of 0.82, close to one standard deviation, and small-group learning had an estimated effect size of 0.49, slightly above the average of 0.4 for the 150 influences, a smaller effect size of 0.15 was estimated for mentoring (Hattie, 2011, p. 251). If drawing nothing more from these numbers than the possibility of some teaching methods being more effective than others, it illustrates the difficulty of knowing how the components in educational interventions contributes to or counteracts a whole. This can be exemplified with the studies by Steckelberg et al. (2009) and Derry et al. (1998), which both used small group work, lectures and class discussions, but only the study by Derry et al. used teachers as mentors and models. Even though the comparability of the results is limited, the relative difference between the intervention and control group was 59% for Derry et al. and 23% for Steckelberg et al. Such numbers show little compliance with the distribution of estimates presented by Hattie, and necessitates another approach to interpreting the results, so while it may enhance the curriculum, it complicates synthesising (Reed et al., 2005).

5.2.3 The study designs

Of the five included studies one was a RCT (Hendricks, 2001 ; Hill, 1998), one was a non-randomised group study with pre- and posttests (Derry et al., 1998), one was a clustered-randomised controlled trial with pre- and posttests (Kaelin et al., 2007) and two were non-randomised group studies with posttests only (Leshowitz et al., 1993 ; Steckelberg et al., 2009). These differing designs contributed to the challenges associated with quantitatively synthesising the individual study results. Generally it is advised that reviewers in educational research should limit their focus to studies using similar designs (Reed et al., 2005), but the handful of studies included in this systematic review, even when including several different kinds of controlled studies, shows that such a limitation in some cases can be viewed as a luxury.

The study by Kaelin et al. (2007) used a clustered design. When using this kind of study design larger groups of people are allocated to interventions instead of individuals, i.e. clusters can be districts, schools or classes. When dealing with clusters as the unit of allocation variability both between groups and within groups can occur and affect the results of the study. Using an intraclass correlation value (ICC) the investigators or the reviewers can adjust for this similarity of individuals within groups. By identifying the ICC either during the study or in retrospect from similar studies, it is possible to adjust the sample size and thereby attain a more realistic effect estimate for the intervention. This has not been done for the study by Kaelin et al. as of yet, but it is planned before publication of the systematic review. After contacting the study authors and establishing that no correction for intraclass correlation had been done in the study phase, an appropriate value was identified in an article by Hedged and Hedberg (2007). On the basis of en extensive amount of data collected from a large amount of school achievement tests in the USA, they have estimated an average ICC of 0.22. By incorporating this value as recommended and exemplified in the Cochrane Handbook (Higgins, Deeks & Altman, 2011a), a more accurate effect estimate can be obtained.

Among the secondary outcomes described in the systematic review attitudes concerning the usefulness of critical appraisal and satisfaction with the educational intervention were specified as relevant. None of these outcomes have been reported in the systematic review. Even when studies as the one by Steckelberg et al. (2009) reports on such outcomes, they are in the form of written comments only supplied by the intervention group. Additionally, the control groups in most cases received instruction as usual, thus making any comparison between the groups futile, should they have given any such statements.

5.2.4 The confounders

As described earlier, teachers represent one of the several sources of confounding factors. Had the included studies to a greater extent reported teacher characteristics and scrutinised the delivery of the interventions more extensively, the lesser an unknown element would the teachers have been. At the same time it may be difficult to estimate the degree of teacher confounding, given that confounders generally correlate with the outcomes in a non-causal way and that the effects in the included studies may not originate solely in the intervention itself, but also in the teacher. As described by Patrício and vaz Carneiro (2012) this challenge lies in the holistic nature of the educational phenomenon that dissociating the causal factor from the confounding factors presents difficulties especially in educational research.

While the study design has the ability to account for several types of confounding factors, some are more difficult to control and should be taken into account by the researchers. Ewert (2009) described three different typed of confounding variables: precursor, concomitant and postexperience.

Precursor variables being the ones brought into the experiment such as age, ethnicity, prior knowledge and gender. All of these have been identified as confounding factors in the included studies to a varying extent. In all the included studies age, gender, socioeconomic status, ethnicity and academic achievement were identified, but only in two studies did the authors recognise the possibility of some of these confounding factors, mainly gender and academic achievement (Kaelin et al., 2007 ; Leshowitz et al., 1993), and only in one was the data analysis adjusted accordingly (Kaelin et al., 2007).

Concomitant variable occur during the experiment or immediately after and comprise group dynamics and events transpiring, it may be course length, activities, teachers or such. While each of the studies described the educational interventions, although to varying degrees, none of the studies gave particularly detailed information on the characteristics of the teachers.

The last category, post experience variables, follows the completion of the intervention and can be exemplified with social desirability, which is the responding to questionnaires with what is presumed to be the desirable answer, or post experience euphoria, which is when the sense of accomplishment obscures the true feelings about one's abilities. No such confounding variables were reported in any of the included studies.

Further, other types of confounding variables can be introduced in the classroom setting as presented by Levin (1992). Time-of-day effects occur when the intervention is delivered to the same participants in the same class period each time, i.e. if the intervention is

based on an intervention delivered in the first class period each Monday morning, the state of the students may affect the results. The "John Henry" effect, named after an American legend steelworker surpassed by a steam drill, is yet another example of confounding variables. This one occurring when the control group is aware of their status and work harder to outdo the intervention group.

Every field of research is vulnerable for confounding variables. Nevertheless, considering the literature on educational research an abundance of variables needs to be considered in this particular field. In the systematic review the most apparent variables have been taken into consideration, but it cannot be denied that some relevant confounding variables have been overlooked.

5.2.5 The heterogeneity

As demonstrated in the previous sections heterogeneity characterises the included studies. When such significant elements as interventions, outcomes and study designs differ to the extent of the included studies, they lead to difficulty in synthesising the results. Nevertheless, while the heterogeneity does impose some limitations, it also offers advantages. It allows the reviewer to assess the generalizability of the interventions across studies and populations as well as the effectiveness and feasibility of educational approaches (Reed et al., 2005).

Keeping the advantages and possibilities of heterogeneity in mind, a call for more research has frequently been expressed in the educational research literature. Some call for more funding to support scientific research on literacy-based health education interventions with the aim of enhancing students' critical health literacy (Deal & Hodges, 2009), while others more generally states the need for high quality, experimental educational research to examine the effectiveness of educational interventions (Slavin, 2002 ; Patricio & vaz Carneiro, 2012). Andrews (2005) further specifies the need for longitudinal studies or studies with delayed posttests, since the use of posttests in the first weeks following the experimental period is likely to show a positive effect for the intervention group. This is of particular relevance in the context of this review, where every outcome was measured soon after the end of the intervention. Furthermore, he stresses not only the need for more evidence on effectiveness, but the need for a stronger methodological focus in future research enhancing the quality of the studies. Such a quality improvement could consist of any methodological improvement from a more unified use of study designs to the use of objective outcome

measures, which is of particular importance since the researcher in many cases is both developer and evaluator of the intervention (Reed et al., 2005).

5.3 Methodological limitations in the systematic review

Since this systematic review was developed as a masters' thesis, some limitations apply. Even though the processes demanding two reviewers in all cases but one were done by two, the comprehensiveness of undertaking a systematic review as a masters' thesis did result in the limitations described underneath.

5.3.1 The protocol

The Cochrane Handbook states that a protocol should be published prior to the initiation of the systematic review to reduce the potential risk of bias in the review process (Green & Higgins, 2011a).

A project plan was developed in advance to conducting the systematic review, outlining the research question, literature search, eligibility criteria, expected data analysis and so on, but the progress of review entailed minor changes and adjustments to the project plan underway. Hence the protocol was published in PROSPERO later than what would be considered ideal (Gundersen et al., 2015). The deviations and changes made from the original project plan have not been judged as substantial, however, in which way it may inadvertently have affected the review is not possible to say.

5.3.2 The literature search

The literature search was rather extensive, comprising both sensitive searches and a variety of sources. Hopefully this will have contributed to minimising the chance of missing relevant studies, but yet another aspect that could have been included in the search was hand searching of relevant journals. Since no database covers either health-related education or medical education, the identification of core journals in the field and a more thorough examination of these journals could have contributed to the exhaustiveness of the search by possibly identifying poorly indexed or non-indexed studies.

The extensiveness of the number of identified references from the literature search did initially seem quite high. Several tests were made with the proximity operators to enhance precision by minimising the number of irrelevant references appearing among the search

results. For every adjustment to the search strategy, a list of the references that would be missed was produced and the first 100-200 references would be perused to ensure no unwanted limitation were done. After these search strategy deliberations, the final amount of identified references was 17.362 unique references. Even though this amount is vast, when compared to two of the more closely related Cochrane reviews by Horsley et al. (2011) and Car (2011) identifying respectively 11.057 and 41.225 references, the amount of references found in this systematic review was judged to be somewhat reasonable. This is also in compliance with the recommendations in the Cochrane Handbook (Lefebvre, Manheimer & Glanville, 2011a) encouraging review authors to aim for sensitivity and accept low precision.

When conducting a literature search, especially a search intended for a systematic review, it is advised that the search strategy is peer reviewed (Sampson et al., 2009). The initial MEDLINE search strategy was reviewed by the information specialist from the EPPI-centre. In regards to the final search strategy the comprehensiveness entailed that at full peer review would be extremely time consuming, so based on the available resources only the final search strategies for the MEDLINE and ERIC databases were peer reviewed. One of the reasons of peer reviewing is to eliminate small mistakes like spelling errors, incorrect combination of search lines and such. To correct for this in some measure, the search strategies for the remaining databases were doubly checked a few days apart before the search was finalised and the results exported to EndNote.

The process of conducting the systematic review was time consuming. Both on account of the large number of references and the methodological challenges along the way, e.g. tool development. Therefore, the time that passed since the major part of the literature search was performed was over a year. Considering that the studies included in this systematic review was published in 1993 (Leshowitz et al.), 1998 (Derry et al.), 1998/2001 (Hill ; Hendricks), 2007 (Kaelin et al.), and 2009 (Steckelberg et al.), the frequency of publication was few and far apart, thus there is not a high likelihood of new studies having been published during the last year. While no definite limit exists for how up-to-date a literature search should be, one of the strengths of a good systematic review is giving an updated summary of available evidence. The Norwegian Knowledge Centre for the Health Services (2013, p. 33) recommends that no more than six to eight months should pass between the literature search is performed and the review is published. This period of time corresponds with the findings in a study by Beller et al. (2013), where 300 systematic reviews were analysed and showed a median of 8 months between last search and publication. For this systematic review the search

has not yet been updated, but a re-running of the literature search is planned closer to publication to ensure that no new and relevant studies remain unidentified.

No language restrictions were made during the literature search, which results in two Chinese and one Arabic article being identified. Judged by the English abstracts none of these articles were eligible for inclusion, but based on similar experiences with other studies, information only available in the full text could change this by revealing relevant characteristics not included in the abstracts. According to Hammick, Dornan and Steinert (2010), the translation of foreign language studies should balance the potential positive benefits against the time and resources needed to conduct the translation. The resources available for this masters' thesis did not include the possibility of translating the studies and since it was not expected that the translation wound lead to inclusion, it was assessed that the exclusion most likely did not influence the results of the review, but the chance remains that it could have.

5.3.3 The initial screening

When the literature search was completed it was very clear that many of the identified references were not in the slightest relevant for inclusion, just as expected for a sensitive search-approach for a review overlapping both health and education thematically. To use the available resources most effectively only one person preformed an initial screening eliminating approximately 55% of the identified studies. In any case of doubt regarding eligibility, the study was included for double screening in the next round of study selection. Even though accuracy was sought during the process, over 9000 references were excluded and erroneous exclusions could have occurred. To control for this a sample could have been drawn from the excluded references and checked by another reviewer. Since this was not done it represents a weakness in the review process, but by both including citation searches and reference list check for the included studies it is hoped that these actions to some extent reduced the risk of bias of missing relevant studies.

5.3.4 Piloting and educational research

To ensure that eligibility criteria are understood in the same manner by each person assessing inclusion to a review a pilot of the study selection process is advisable. For this review piloting was not done as a single test, but the number of references made study selection a

process which of practical reasons was divided into several sessions. In that way a piloting test of the criteria was performed, but integrated in the study selection.

Had a pilot test been performed initially as a separate process, it could possibly have refined the study selection and calibrated the reviewers to a higher degree. Additionally a kappa score could have been calculated to assess the agreement between the reviewers. The Cochrane Handbook mentions the use of this measure especially early in the review process, but they do not necessarily recommend it invariably. Therefore, based on the scope of the review and the overall level of agreement between the reviewers this was not done.

One of the explanations for discrepancies in study selection between the two reviewers was quite possibly the lack of experience with educational research in one of the reviewers. While this may have led to the initial inclusion of more irrelevant studies than else, there was no reason to believe that relevant references were excluded because of this. Even though the study selection process was not organised with this specific goal in mind, the Cochrane Handbook (Higgins & Deeks, 2011) expresses the advantages of including both reviewers with and without knowledge of the content in the selection process, since such a division may reduce the influence of pre-formed opinions on the study selection.

5.3.5 Presentation of the results

Had it been reasonable the results would have been synthesised in one or more meta-analyses. As with many other systematic reviews on effects of interventions, the opportunity to give a synthesised effect estimate based on several studies would have been ideal. This may be of particular interest in reviews incorporating non-randomised studies, since it is presumed that the biases present in the studies would be averaged out by the synthesis, giving a more reliable effect estimate (Colliver, Kucera & Verhulst, 2008). Nevertheless, because of both the practical and methodological heterogeneity this was not feasible. Instead the results were presented narratively for each study and similarities were described.

A narrative summarisation can connect the studies and describe similarities and dissimilarities, but in comparison to a meta-analysis this approach reintroduces much of the bias sought minimised through the earlier stages of the systematic review (Andrews, 2005). When a reviewer moves beyond a mere presentation of more or less objectively calculated effect estimates and begins to describe the results, objectivity logically decreases. Even so Colliver, Kucera and Verhulst (2008) tentatively concludes that results from quasi-experimental studies may actually be more correctly presented through systematic narrative

reviews. While stressing the need for further research on the topic, they state that constant biases and confounders shared by a group of studies, such as quasi-randomised trials, may undermine the meta-analysis by threatening the validity of the results. Although no such claims were taken into considerations when developing this systematic review, it corroborates the contribution to research a review of this nature can provide.

For further presentation of the results a summary of findings table was developed in GRADE. When both the outcomes included in this review and the outcomes described in the studies were taken into account, the standard GRADE-table only partly allowed for the desired elements to be expressed. Because of this, the GRADE-approach was used as expected to express the quality of evidence itself, but the table was slightly modified to present the findings in the most suitable way.

5.4 Findings

Looking at the findings of this systematic review in the light of the presented limitations imposed both by the included studies and the methodological limitation of the review process itself, no firm conclusions could be drawn based on the available evidence. The results consistently indicated a beneficial effect of the educational interventions, but the findings came from research open to many types of bias and were in several cases only marginally statistically significant. When this was combined with the limitations in generalizability, the result was little confidence in the effect estimates. The results should be interpreted with caution.

5.4.1 Relation to other studies

The dearth of similar systematic reviews on educational interventions for enhancing critical appraisal abilities limits the possibility of comparing results. In the systematic review the results has been compared to some of the most closely related systematic reviews. They, like this review, generally present few studies of varying quality, with small effects and relatively inconclusive findings. Although one element seems to present itself in each review, the indication of positive effects on learning outcomes.

6. Conclusion

6.1 Implications for practice

Only five studies were identified, they all had high or moderate risk of bias and the quality of evidence was for each study assessed as low. These facts alone entails that no firm conclusions can be drawn. Each of the studies did however show positive effects of the interventions which can indicate that educational interventions may enhance the critical appraisal abilities of health-related claims in adolescents, but across all studies only small effects were demonstrated, so caution should be shown when interpreting the findings.

6.2 Implications for research

The findings in this systematic review were greatly influenced by the heterogeneity in the included studies. As mentioned earlier, this heterogeneity characterises educational research in general, not just the specific topic dealt with in this review. Future research would greatly benefit from the use of more methodologically rigid, experimental studies using both objective outcome measures and delayed posttests. Conducting more studies using similar methods aimed at minimising bias will allow for synthesising of the evidence and thereby building a cumulative knowledge on which educational interventions that has the most effect. Consequently the conclusion is not merely that more research is needed, but that more research of a higher quality is needed.

7. References

Andrews, R. (2005) The Place of Systematic Reviews in Education Research *British Journal of Educational Studies,* 53 (4), s. 399-416.

Armstrong, R., Hall, B. J., Doyle, J. & Waters, E. (2011) 'Scoping the scope' of a cochrane review. *Journal of Public Health,* 33 (1), s. 147-150.

Belland, B. R., Glazewski, K. D. & Richardson, J. C. (2011) Problem-Based Learning and Argumentation: Testing a Scaffolding Framework to Support Middle School Students' Creation of Evidence-Based Arguments. *Instructional Science: An International Journal of the Learning Sciences,* 39 (5), s. 667-694.

Beller, E. M., Chen, J. K., Wang, U. L. & Glasziou, P. P. (2013) Are systematic reviews up-to-date at the time of publication? *Systematic Reviews,* 2, 36.

Bennett, Judith, Lubben, F., Hogarth, S. & Campbell, B. (2005) Systematic reviews of research in science education: rigour or rigidity? *International Journal of Science Education,* 27 (4), s. 387-406.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956) *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* New York, David McKay Company.

Brown, S. L., Teufel, J. A. & Birch, D. A. (2007) Early adolescents perceptions of health and health literacy. *The Journal of school health,* 77 (1), s. 7-15.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D. & Tyrer, P. (2000) Framework for design and evaluation of complex interventions to improve health. *BMJ,* 321:694.

Car, J., Lang, B., Colledge, A., Ung, C. & Majeed, A. (2011) Interventions for enhancing consumers' online health literacy. *The Cochrane database of systematic reviews*,15 (6), CD007092.

Cochrane Consumer and Communication Review Group (2013) *Data extraction template for Cochrane reviews (updated February 2013).* Available from: http://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/uploads/DET_2013_0.doc [Read May 14th 2015]

Colliver, J. A., Kucera, K. & Verhulst, S. J. (2008) Meta-analysis of quasi-experimental research: are systematic narrative reviews indicated? *Medical Education*, 42 (9), s. 858-865.

Cooper, B. E. J., Lee, W. E., Goldacre, B. M. & Sanders, T. A. B. (2012) The quality of the evidence for dietary advice given in UK national newspapers. *Public Understanding of Science*, 21 (6), s. 664-673.

Davies, K. S. (2011) Formulating the Evidence Based Practice Question: A Review of the Frameworks. *Evidence Based Library and Information Practice*, 6 (2), s. 75-80.

Deal, T. B. & Hodges, B. (2009) *Role of 21st Century Schools in Promoting Health Literacy*. Washington D.C., NEA Health Information Network.

Deeks, J. J., Higgins, J. P. T. & Altman, D. G. (2011) Analysing data and undertaking meta-analyses. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap. 9.

Derry, S. J., Levin, J. R., Osana, H. P. & Jones, M. S. (1998) Developing middle-school students' statistical reasoning abilities through simulation gaming. I: Lajoie S. P. red. *Reflections on statistics: Learning, teaching, and assessment in Grades K-12.* Mahwah, Lawrence Erlbaum Associates Publishers, s. 175-195.

Effective Practice and Organisation of Care (EPOC) (2015) *Which studies should be included in the Characteristics of excluded studies table? EPOC Resources for review authors* [Internet], Oslo, Norwegian Knowledge Centre for the Health Services,. Available from: http://epoc.cochrane.org/epoc-specific-resources-review-authors [Read May 7th 2015].

Egger, M. & Smith, G. D. (1998) Meta-analysis bias in location and selection of studies. *BMJ,* 316 (7124), s. 61-66.

Eva, K. (2008) On the limits of systematicity. *Medical education,* 42 (9), s. 852-853.

Ewert, A. (2009) Creating Outcomes through Experiential Education: The Challenge of Confounding Variables. *The Journal of experiential education,* 31 (3), s. 376-389.

Gegner, J., Mackay, D. & Mayer, R. (2009) Computer-supported aids to making sense of scientific articles: cognitive, motivational, and attitudinal effects. *Educational Technology Research and Development,* 57 (1), s. 79-97.

Glenton, C., Paulsen, E. J. & Oxman, A. D. (2005) Portals to Wonderland: health portals lead to confusing information about the effects of health care. *BMC Med Inform Decis Mak,* 5, s. 7.

Green, S. & Higgins, J. P. (2011a) Preparing a Cochrane review. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap. 2.

Green, S. & Higgins, J. P. (2011b) What is a systematic review? I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration., chap. 1.2.2.

Gundersen, M. W., Nordheim, L., Espehaug, B. & Flottorp, S. (2015) *Effects of educational interventions for enhancing adolescents abilities in critical appraisal of health claims: a systematic review*. Prospero: International Prospective Register of Systematic Reviews. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015017936 [Read May 5th 2015]

Haig, A. & Dozier, M. (2003) BEME Guide no 3: systematic searching for evidence in medical education--Part 1: Sources of information. *Medical teacher,* 25 (4), s. 352-63.

Hammick, M., Dornan, T. & Steinert, Y. (2010) Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13. *Medical teacher*, 32 (1), s. 3-15.

Hargreaves, D. (1996) Teaching as a research-based profession: possibilities and prospects. I: *The Teacher Training Agency Annual Lecture, London.*

Hatlevik, O. E., Egeberg, G., Guðmundsdóttir, G. B., Loftsgarden, M. & Loi, M. (2013) Monitor skole 2013: Om digital kompetanse og erfaringer med bruk av IKT i skolen. [Oslo], Senter for IKT i utdanningen.

Hattie, J. (2011) *Visible learning for teachers : maximizing impact on learning.* Hoboken, N.J, Taylor & Francis.

Hedges, L. V. & Hedberg, E. C. (2007) Intraclass Correlation Values for Planning Group-Randomized Trials in Education. *Educational Evaluation and Policy Analysis*, 29 (1), s. 60-87.

Hendricks, C. C. (2001) Teaching causal reasoning through cognitive apprenticeship: What are results from situated learning? *The Journal of Educational Research,* 94 (5), s. 302-311.

Higgins, J. P. T., Altman, D. G. & Sterne, J. A. C. (2011) Assessing risk of bias in included studies. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap. 8.

Higgins, J. P. T. & Deeks, J. J. (2011) Selecting studies and collecting data. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap. 7.

Higgins, J. P. T., Deeks, J. J. & Altman, D. G. (2011a) Cluster-randomized trials. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap. 16.3.

Higgins, J. P. T., Deeks, J. J. & Altman, D. G. (2011b) Special topics in statistics. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap. 16.

Higgins, J. P. T. & Green, S. (2011) *Cochrane handbook for systematic reviews of interventions.* Version 5.1.0 [updated March 2011]. [S.l.], The Cochrane Collaboration.

Higgins, J. W., Begoray, D. & MacDonald, M. (2009) A social ecological conceptual framework for understanding adolescent health literacy in the health education classroom. *American Journal of Community Psychology*, 44 (3), s. 350-362.

Hill, C. C. (1998) The effects of situated learning, abstracted instruction, and teaching for transfer on students' use of statistical reasoning to solve real-world problems. *Dissertation Abstracts International Section A: Humanities and Social Sciences,* 59 (5-A), s. 1457.

Horsley, T., Hyde, C., Santesso, N., Parkes, J., Milne, R. & Stewart, R. (2011) Teaching critical appraisal skills in healthcare settings. *Cochrane Database of Systematic Reviews,* 9(11), CD001270.

Jenuwine, E. S. & Floyd, J. A. (2004) Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. *Journal of the Medical Library Association : JMLA,* 92 (3), s. 349-53.

Joint Committee on National Health Education Standards (2007) *National Health Education Standards*. Available from: http://www.cdc.gov/healthyyouth/sher/standards/3.htm [Read May 4th 2015]

Kaelin, M. A., Huebner, W. W., Nicolich, M. J. & Kimbrough, M. L. (2007) Field Test of an Epidemiology Curriculum for Middle School Students. *American journal of health education*, 38 (1), s. 16-31.

Kirkpatrick, D. L. (1967) Evaluation of training. I: Craig, L. & Bittel, L. R. red. *Training and Development Handbook.* New York, McGraw-Hill, s. 87-112.

Kjærnsli, M. & Olsen, R. V. (2013) *Fortsatt en vei å gå : norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012.* Oslo, Universitetsforl.

Krathwohl, D. R. (2002) A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice,* 41 (4), s. 212-218.

Laugksch, R. C. (2000) Scientific literacy: A conceptual overview. *Science Education,* 84 (1), s. 71-94.

Lefebvre, C., Manheimer, E. & Glanville, J. (2011a) Searching for studies. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap. 6.

Lefebvre, C., Manheimer, E. & Glanville, J. (2011b) Structure of a search strategy. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap. 6.4.2.

Leshowitz, B., Jenkens, K., Heaton, S. & Bough, T. L. (1993) Fostering critical thinking skills in students with learning disabilities: an instructional program. *Journal of Learning Disabilities,* 26 (7), s. 483-90.

Levin, J. R. (1992) On Research in Classrooms. *Mid-Western Educational Researcher,* 5 (1), s. 2-6.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J. & Moher, D. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ,* 339: b2700.

Manganello, J. A. (2008) Health literacy and adolescents: a framework and agenda for future research. *Health Education Research,* 23 (5), s. 840-7.

McMaster University & Evidence Prime Inc. (2015) GRADEpro Guideline Development Tool. Available from: http://www.guidelinedevelopment.org/ [Read 12.05.2015]

Ministry of Education and Research (2004) *Kultur for læring.* [Oslo], [Regjeringen].

Ministry of Health and Care Services (2012) *God kvalitet - trygge tjenester: kvalitet og pasientsikkerhet i helse- og omsorgstjenesten.* [Oslo], [Regjeringen].

Nabulsi, M., Harris, J., Letelier, L., Ramos, K., Hopayian, K., Parkin, C., Porzsolt, F., Sestini, P., Slavin, M. & Summerskill, W. (2007) Effectiveness of education in evidence-based healthcare: the current state of outcome assessments and a framework for future evaluations. *International journal of evidence-based healthcare,* 5 (4), s. 468-476.

Nielsen-Bohlman, L., Panzer, A. M., Kindig, D. A. & Institute of Medicine Committee on Health, L. (2004) Health literacy : a prescription to end confusion. Washington, D.C, National Academies Press.

Norway. Patients' Rights Act (1999) *The Act of 2 July 1999 No. 63 relating to Patients' Rights*. Available from: http://www.lovdata.no/all/nl-19990702-063.html [Read May 5th 2015].

Norwegian Directorate for Education and Training (2013) *Natural Science subject curriculum - Competence aims - competence aims after Year Level 10*. Available from: http://www.udir.no/kl06/NAT1-03/Kompetansemaal/?arst=98844765&kmsn=-1974299133 [Read May 5th 2015].

O'Connor, D., Green, S. & Higgins, J. P. T. (2011a) Defining the review question and developing criteria for including studies. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* The Cochrane Collaboration, chap. 5.

O'Connor, D., Green, S. & Higgins, J. P. T. (2011b) Defining types of study. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* The Cochrane Collaboration, chap. 5.5.

OECD (2012) *Programme for International Student Assessment (PISA). Results from PISA 2012: Norway*, Paris, OECD,. Available from: http://www.oecd.org/pisa/keyfindings/PISA-2012-results-norway.pdf [Read May 6th 2015].

OECD (2013) *PISA 2015. Draft Science Framework*. Paris, Oecd Publications. Available from: http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Science%20Framework%20.pdf [Read May 6th 2015]

OECD (2015) *About PISA* [Internett], Paris, OECD,. Available from: http://www.oecd.org/pisa/aboutpisa/ [Read May 6th 2015].

Olson, D. R. (2004) The Triumph of Hope Over Experience in the Search for "What Works": A Response to Slavin. *Educational Researcher,* 33 (1), s. 24-26.

Patricio, M. & vaz Carneiro, A. (2012) Systematic reviews of evidence in medical education and clinical medicine: is the nature of evidence similar? *Medical teacher,* 34 (6), s. 474-82.

Pettersen, S. (2001) Undervisning for naturvitenskapelig allmenndannelse: vektlegging av vitenskapens innhold eller prosesser? *Norsk pedagogisk tidsskrift,* 1, s. 49-59.

Pettersen, S. (2003) Er også naturfagdidaktikk godt for helsen? I: Bungum, B., Jorde, D. & Sjøberg, S. red. *Naturfagdidaktikk : perspektiver, forskning, utvikling.* Oslo, Gyldendal akademisk, s. s. 373-388.

Reed, D., Price, E. G., Windish, D. M., Wright, S. M., Gozu, A., Hsu, E. B., Beach, M. C., Kern, D. & Bass, E. B. (2005) Challenges in systematic reviews of educational intervention studies. *Annals of Internal Medicine,* 142 (12 Pt 2), s. 1080-9.

Reeves, B. C., Deeks, J. J., Higgins, J. P. T. & Wells, G. A. (2011) Including non-randomized studies. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, kap 13.

Ryan, R., Hill, S., Prictor, M. & McKenzie, J. (2013) *Cochrane Consumers and Communication Review Group. Study Quality Guide*. Available from: http://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/uploads/StudyQualityGuide_May%202013.pdf [Read May 2th 2015]

Ryder, J. (2001) Identifying Science Understanding for Functional Scientific Literacy. *Studies in Science Education*, 36 (1), s. 1-44.

Ryder, J. (2002) School science education for citizenship: Strategies for teaching about the epistemology of science. *Journal of Curriculum Studies*, 34 (6), s. 637-658.

Sampson, M., McGowan, J., Cogo, E., Grimshaw, J., Moher, D. & Lefebvre, C. (2009) An evidence-based practice guideline for the peer review of electronic search strategies. *Journal of Clinical Epidemiology*, 62 (9), s. 944-52.

Schünemann, H. J., Oxman, A. D., Vist, G. E., Higgins, J. P. T., Deeks, J. J., Glasziou, P. P. & Guyatt, G. H. (2011) Interpreting results and drawing conclusions. I: Higgins, J. P. & Green, S. red. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, chap 12.

Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002) *Experimental and quasi-experimental designs for generalized causal inference.* Boston, Houghton Mifflin.

Slavin, R. (2002) Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational researcher*, 31 (7), s. 15-21.

Sorensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z. & Brand, H. (2012) Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health*, 12, s. 80.

Squires, J. E., Valentine, J. C. & Grimshaw, J. M. (2013) Systematic reviews of complex interventions: framing the review question. *Journal of Clinical Epidemiology,* 66 (11), s. 1215-22.

Steckelberg, A., Hülfenhaus, C., Kasper, J. & Mühlhauser, I. (2009) Ebm@school--a curriculum of critical health literacy for secondary school students: results of a pilot study. *International journal of public health*, *54* (3), s. 158-65.

Sterne, J., Higgins, J., Reeves, B. & on behalf of the development group for ACROBAT-NRSI (2014) *A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT- NRSI), Version 1.0.0* [Internett]. Available from: http://www.riskofbias.info [Read May 5th 2015].

The Campbell Collaboration (2004) *Research Design Policy Brief* http://www.campbellcollaboration.org/artman2/uploads/1/C2_Research_Design_Policy_Brief-2.pdf [Read April 22nd 2015]

The Cochrane Collaboration (2014) Review Manager (RevMan). Version 5.3 utg. The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen.

The Evidence for Policy and Practice Information and Co-ordinating Centre (2009) *History of systematic reviews* [Internett]. Available from: http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=68 [Read May 7th 2015].

The Norwegian Knowledge Centre for the Health Services (2013) Slik oppsummerer vi forskning : håndbok for Nasjonalt kunnskapssenter for helsetjenesten. 3.2. rev. utg. Oslo, Nasjonalt kunnskapssenter for helsetjenesten.

The Reference and Web Services Section (2015) *Health Literacy Search* [Internett], Bethesda, National Library of Medicine. Available from: <http://www.nlm.nih.gov/services/queries/health_literacy.html> [Read February 14th 2015].

Torgerson, D. J. P. & Torgerson, C. J. D. (2008) *Designing Randomised Trials in Health, Education and the Social Sciences : An Introduction.* Basingstoke, Palgrave Macmillan.

World Health Organization (2013) *Adolescent health* [Internett]. Available from: <http://www.who.int/topics/adolescent_health/en/> [Read March 2nd 2015].

Appendix I

Data extraction form - template

SOURCE

RefID (first author, year):	
Citation:	Author AB, Author CD, Author EF. (2014) This is the article. *Journal of journals* 1 (1), s. 1-2
Country:	
Publication type:	☐ Journal article ☐ Book chapter ☐ Report ☐ Book ☐ Dissertation ☐ Other (specify)_____
Date of completing form:	(finished)

METHOD

Aim of study:	
Study design	☐ Randomised controlled study ☐ Non-randomised controlled study ☐ Interrupted time series ☐ Pre-test and post-test ☐ Pre-test and post-test ☐ Post-test only ☐ Post-test only Comments:_____
Unit of allocation	☐ Students ☐ Classes ☐ Teachers ☐ Schools ☐ Other, specify:
Type of control:	☐ Traditional/standard teaching ☐ Other intervention
No. schools in study:	
Year(s) of data collection:	

PARTICIPANTS – STUDENTS

	All	I_1	I_2	C
Eligibility criteria (enter in appropriate column if criteria differ by group):				
Number identified:				
Attrition:				
Number included:				
All accounted for?	☐ Yes ☐ No			
Mean age (SD):				
Males (percentage):				
Grade level:				
Ethnicity:				
Academic achievement:				
SES:				
Other characteristics:				

PARTICIPANTS – EDUCATION PROVIDERS

	All	I_1	I_2	C
Eligibility criteria(enter in appropriate column if criteria differ by group):				
Number identified:				
Attrition:				
Number included:				
All accounted for?	☐ Yes ☐ No			
Mean age (SD):				
Occupation:				
Years of experience:				
Field:				
Males(percentage):				
Other characteristics:				

CONFOUNDERS

Tick if considered to be a confounder. Tick last column to indicate whether groups were considered different by the researchers.

STUDENTS:	**Confounder?**	**Different?**	**Comments**
Age:	☐	☐	
Gender:	☐	☐	
Grade level:	☐	☐	
Ethnicity:	☐	☐	
Academic achievement:	☐	☐	
SES:	☐	☐	
Other:	☒		
	☐	☐	
	☐	☐	
	☐	☐	
EDUCATION PROVIDERS:			
Age:	☐	☐	
Teaching experience:	☐	☐	
Teaching styles:	☐	☐	
Other:			
	☐	☐	
	☐	☐	
	☐	☐	

INTERVENTION AND COMPARISON CHARACTERISTICS

	I_1	I_2	C
Theory:			
Learning objectives			
Teaching method:			
Contents:			
Number of lessons:			
Duration of lessons *(min)*:			
Duration of intervention *(pre-test/start of intervention to post-test in days)*:			
Length of follow-up (days after post-test):			
Compliance/Adherence (percent):			

CONTINUOUS OUTCOMES

Outcome: _____

Type of outcome:
□ Cognitive
□ Behavioural

Page number or table number in article: ___

Measurement method (questionnaire, interview etc.)**:**
_____

Instrument (Name of method/instrument used)**:**
_____

Timing of outcome assessment:
Post-test: _____
Follow-up:_____

Effect estimate: _____

P-value: _____

	Baseline			**Baseline excluding drop-outs**			**Post-test**			**Follow-up**		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Intervention 1												
Intervention 2												
Comparison												
Comments												

Effect estimates if non-randomised controlled study:

	UNADJUSTED ESTMATES				**ADJUSTED ESTIMATES**		
	Baseline (number identified)	**Number analysed**	**Effect estimate and SD**	**Baseline (number identified)**	**Number analysed**	**Effect estimate and SD**	**Confounders included in adjusted analysis**
Intervention 1			Mean:			Mean:	
			SD:			SD	
Intervention 2			Mean:			Mean:	
			SD:			SD	
Comparison			Mean:			Mean:	
			SD:			SD:	
Comments							

Outcome: _____

Type of outcome:
□ Cognitive
□ Behavioural

Page number or table number in article:

Measurement method (questionnaire, interview etc.)**:**
_____

Instrument (Name of method/instrument used)**:**
_____

Timing of outcome assessment:
Post-test: _____
Follow-up:_____

Effect estimate: _____

P-value: _____

DICHOTOMOUS OUTCOMES

	Baseline			**Baseline excluding drop-outs**			**Post-test**			**Follow-up**		
	N	n	Percent	N	n	Percent	N	n	Percent	N	n	Percent
Intervention 1												
Intervention 2												
Comparison												
Comments												

n = observed

Effect estimates if non-randomised controlled study:

	UNADJUSTED ESTMATES			**ADJUSTED ESTIMATES**			
	Baseline (number identified)	**Number analysed**	**Effect estimate and SE / CI**	**Baseline (number identified)**	**Number analysed**	**Effect estimate and SE / CI**	**Confounders included in adjusted analysis**
Intervention 1			HR ☐ OR☐ RR ☐			HR ☐ OR☐ RR ☐	
			SE ☐ CI☐			SE ☐ CI☐	
Intervention 2			HR ☐ OR☐ RR ☐			HR ☐ OR☐ RR ☐	
			SE ☐ CI☐			SE ☐ CI☐	
Comparison			HR ☐ OR☐ RR ☐			HR ☐ OR☐ RR ☐	
			SE ☐ CI☐			SE ☐ CI☐	
Comments							

Appendix II

Risk of bias assessment tool - template

StudyID (first author, year):					
Review author (intials):					
Citation:	Author AB, Author CD, Author EF. (2014) This is the article. *Journal of journals* 1 (1), s. 1-2				
Study design:	☐ Randomised controlled study ☐ Pre-test and post-test ☐ Post-test only		☐ Non-randomised controlled study ☐ Pre-test and post-test ☐ Post-test only		☐ Interrupted time series Comments:
Unit of allocation	☐ Students	☐ Classes	☐ Teachers	☐ Schools	☐ Other, specify:

	Risk of bias	Criteria	Comment
1 Sequence generation *(assess whether the method used to generate the allocation sequence is sufficient to produce a non-predictable assignment pattern*)	☐ High	Not appropriate, poorly executed	
	☐ Unclear		
	☐ Low	Adequate	
2 Allocation concealment *(assess whether allocation concealment approaches, i.e. sequence for allocating participants to groups, are truly hidden from investigators)*	☐ High	Not appropriate, poorly executed	
	☐ Unclear		
	☐ Low	Adequate	
3 Baseline characteristics and outcome measurements comparable *(comparability of groups, relevant comparisons and measures)*	☐ High	Not measured or significant differences between groups	
	☐ Unclear		
	☐ Low	Similar groups	
4 Blinding of students and education providers *(assess whether students or teachers were blinded from knowledge of which intervention a student received)*	☐ High	No blinding	
	☐ Unclear		
	☐ Low	Adequate blinding	
5 Departures from intended interventions *(assess whether co-interventions were equally distributed, contamination prevented, and implementation failure avoided)*	☐ High	Substantial departures	
	☐ Unclear		
	☐ Low	None or similar across groups	
6 Blinding of outcome assessment *(assess whether outcome assessors were blinded from knowledge of which intervention a student received)*	☐ High	No blinding	
	☐ Unclear		
	☐ Low	Adequate blinding	
7 Incomplete outcome data (*assess whether outcome data for each main outcome is complete)*	☐ High	No report of attrition rate, not taken into account	
	☐ Unclear		
	☐ Low	Reported separately for groups	
8 Selective reporting *(assess if all outcomes are accounted for in results)*	☐ High	Reported on some outcomes	
	☐ Unclear		

	☐ Low	Reported on all outcomes	
9 Outcome measures reliable? *(assess if reliable, appropriate, indirect)*	☐ High	Not validated, questionable	
	☐ Unclear		
	☐ Low	Validated and reliable	
10 Outcome measures validated? *(assess if validated)*	☐ High	Not validated	
	☐ Unclear		
	☐ Low	Validated	
11 Other bias *(assess if measures taken and if avoided)*	☐ High	Important concerns	
	☐ Unclear		
	☐ Low	None	
12 Overall assessment*	☐ High	High risk of bias	Poor quality
	☐ Unclear	Moderate risk of bias	Moderate quality
	☐ Low	Low risk of bias	Good quality

*Summary assessment according to Cochrane: http://handbook.cochrane.org/chapter_8/8_7_summary_assessments_of_risk_of_bias.htm

Domain	**O_1**	**O_2**	**O_3**
Sequence generation			
Allocation concealment			
Baseline characteristics and outcome measurements comparable			
Blinding of students and education providers			
Departures from intended interventions			
Blinding of outcome assessment			
Incomplete outcome data			
Selective reporting			
Outcome measures reliable			
Outcome measures validated			
Other bias			
Overall assessment			

Non-randomised studies: See next page for assessment form for how researchers dealt with confounders.

ASSESSMENT OF HOW RESEARCHERS DEALT WITH CONFOUNDING IN NON-RANDOMISED CONTROLLED STUDIES

Are relevant confounders described?		
☐	YES: Confounders are described and in accordance with pre-specified confounders (*Specify confounders in Data extraction form*)	
☐	NO: No report of confounders in text or tables OR only a subset of relevant confounders are described	
☐	UNCLEAR: Not clear in paper (e.g. statements such as «the groups did not differ with regard to relevant characteristics»)	
Were confounding variables measured validly and reliably?		
☐	YES Describe measures used:	
☐	NO: Confounding variables are measured using measures that neither are reliable nor valid.	
☐	UNCLEAR: No information about the measures used for confounding variables	
Were appropriate methods used to control for confounding at design stage or at analysis stage?		
YES - at design stage:		
☐	Matching	Describe variables on which subjects matched:
☐	Restriction	Describe variables for restriction:
YES - at analysis stage (the variable is not controlled in the design, but rather in the analysis of the data):		
☐	Stratification	Describe prognostic factors (variables) on which subgroups were defined:
☐	Regression	Describe variables used in regression model:
☐	Propensity scores - matching	Describe variables on which propensity score is based:
☐	Propensity scores - regression	Describe variables on which propensity score is based:
☐	Propensity scores - IPW	Describe variables on which propensity score is based:
☐	NO: The authors present uncontrolled effect estimates only	
☐	UNCLEAR: The authors state that the effect estimate is controlled for confounding but provide no information about the methods used in this regard.	

GRADE assessments with reasons

No. of students (studies)	Outcome	Study design	Study limitations	Directness	Consistency	Precision	Reporting	GRADE
Educational interventions that compare different teaching modalities								
220 (1 RCT) (Hendricks, 2001 ; Hill, 1998)	**Knowledge and skills relevant for critical appraisal:** – **Understanding causality**	4 The evidence was by default graded as high as the study was an RCT.	-1 Serious limitations because sequence generation and allocation concealment is not reported	-1 One school only and academic achievement among students is generally low	0 Not relevant (one study)	-1 Only one study	0 Low probability of publication bias. It is more likely that the problem area is unexplored rather than studies have been unpublished due to no effects.	1 Very low
220 (1 RCT) (Hendricks, 2001 ; Hill, 1998)	**Critical appraisal-related outcomes:** – **Causal reasoning**	4 See above	-1 See above	-1 See above	0 See above	-1 See above	0 See above	1 Very low
Educational interventions compared to instruction as usual								
512 (1 cluster-RCT) (Kaelin et al., 2007)	**Knowledge and skills relevant for critical appraisal:** – **Understanding epidemiology**	4 The evidence was by default graded as high as the study was a cluster-RCT.	-2 Very serious limitations because of insufficient sequence generation, allocation concealment not reported, no adjustment for clustering effects, and selective reporting of results	0 No serious indirectness	0 Not relevant (one study)	-1 Only one study.	0 Low probability of publication bias. It is more likely that the problem area is unexplored rather than studies have been unpublished due to no effects.	1 Very low
255 (1 non-RCT) (Steckelberg et al., 2009)	**Knowledge and skills relevant for critical appraisal:** – **Understanding EBM aspects**	2 The evidence was by default graded as low as the study was a non-randomised controlled study.	-2 Very serious limitations due to no adjustment for clustering effects or possibly confounding factors, and unclear or high risk of bias within study assessments in general	-1 Serious indirectness because the educational providers were researchers and not teachers	0 Not relevant (one study)	-1 Only one study	0 Low probability of publication bias. It is more likely that the problem area is unexplored rather than studies have been unpublished due to no effects.	-2 Very low
9 classes (1 non-RCT), 55 students (1 non-RCT) (Derry et al., 1998 ; Leshowitz et al., 1993)	**Critical appraisal-related outcomes:** – **Causal reasoning**	2 The evidence was by default graded as low as both studies were classified as non-randomised controlled studies.	-2 Very serious limitations due to no or insufficient adjustment for possibly confounding factors, and unclear or high risk of bias within and across study assessments in general	-1 Serious indirectness because the educational providers were researchers and not teachers in one study, and the intervention was aimed at students with learning disabilities in the other study. The studies are conducted in one or two schools only	-1 Serious inconsistencies due to heterogeneity in scales	0 High risk of bias due to selective reporting (no SDs and only non-exact p-values in one or both studies) are already considered under "Study limitations". No additional serious imprecisions.	0 Low probability of publication bias. It is more likely that the problem area is unexplored rather than studies have been unpublished due to no effects.	-2 Very low

Article manuscript

Effects of educational interventions for enhancing adolescents' abilities in critical appraisal of health claims: a systematic review

Master's student: Malene Wøhlk Gundersen

Master's program: Master in Evidence-Based Practice for Health and Social Sciences

Faculty: Faculty of Health and Social Sciences

Institution: Bergen University College

Intended journal: PLoS One

Effects of educational interventions for enhancing adolescents' abilities in critical appraisal of health claims: a systematic review

Malene W. Gundersen, Master's student, Centre for Evidence-based Practice, Faculty of Health and Social Sciences, Bergen University College, Norway.

Abstract

Background: The ability to appraise data and evidence scientifically, one of several competencies OECD stresses as a major goal for science education in schools, requires knowledge about science. Enhancement of this ability in adolescence might entail better understanding of how to identify reliable health claims. Until now no secondary research has been conducted on this topic.

Objective: To identify, appraise and synthesize studies examining the effectiveness of school-based educational interventions aiming to enhance adolescents' abilities in critically appraising health claims.

Methods: A systematic review including randomised and non-randomised controlled trials and interrupted time series. To identify these, an extensive literature search was conducted in health-related and educational databases. Two authors independently performed study selection, data extraction and risk of bias assessment with adapted versions of the Cochrane Collaborations tools and resources.

Results: Of over 17000 identified references five studies published in six reports was included: one randomised controlled trial, one cluster-randomised controlled trial and three non-randomised controlled trials. One study compared different teaching modalities, while four studies compared educational interventions to instruction as usual. Risk of bias was predominantly rated as high. The results showed small, but statistically significant effects in favour of the interventions.

Conclusion: The small number of included studies, their heterogeneity and low methodological quality inhibits any firm conclusions on the effects of school-based interventions for enhancing critical appraisal abilities in adolescents. The results indicate a beneficial effect, but must be interpreted with caution.

Keywords: Adolescent (MeSH), Health literacy (MeSH), Education (MeSH), Critical appraisal

Introduction

With the multitude of channels distributing information in the today's digital society developing the health literacy of children and adolescents will allow them to make better judgments concerning the own health both now and in the future. From 1999 to 2009 the time youth aged 8-18 spent on some kind of media increased from 6 hours and 19 minutes to 7 hours and 38 minutes (1). Whether adolescents either actively seek health related information or simply happen upon it, they are major consumers of many types of media, from the Internet to the news media and magazines (2-4). As several studies have stressed the challenge of judging how reliable the information is (4-7), it is credible that the schools can be key institutions in developing students' skills in critically appraising the health related claims and information they encounter (8, 9).

Appraisal skills are crucial to a person's overall health literacy and refer to the ability to interpret, filter, judge and evaluate health information (10). In the medical literature, the term critical appraisal is frequently used to describe the process of evaluating the validity of scientific articles. However, this term could equally apply to the process of scientifically evaluating specific lay health content and health claims in contemporary media (11).

In a survey conducted in Norwegian university colleges of nursing, physiotherapy, social educator, and radiography studies the results indicated that no matter the level of former upper secondary school biology or science courses, the students did not have sufficient knowledge of how to assess the scientific validity in deficient health news briefs (12). The scientific literacy necessary to evaluate such health claims and information requires both critical thinking and some degree of knowledge about science (8, 9).

Even though health claims may be presented in a way appearing to be scientific sound but without actually being so, use of preliminary data, pseudo-scientific facts, poorly executed studies or inflated expert opinions makes it harder to ascertain whether the scientific foundation is reliable, making scientific literacy an essential trait (13, 14).

Notably, achieving scientific literacy involves developing knowledge about science (15, 16). This refers to knowledge about the methods scientists use to obtain valid and precise data, and an understanding of the ways in which claims in science are developed and justified (15). Knowledge about science is central to students' overall scientific literacy and critical thinking abilities to help them decide whether claims in contemporary media are warranted and can be trusted (15-17).

While no easy answer to this question exists, Manganello (8) recognizes the necessity of developing and evaluating school-based interventions to enable adolescents to make well founded health-related decisions by understanding and using health-related information.

An initial scoping search did not succeed in identifying any kind of reviews on school-based interventions to enhance adolescents' abilities critically appraise health claims, though some dealt with relating topics. A systematic review assessed the effect of interventions aimed at enhancing consumers' online health literacy. Positive effects information evaluation skills were identified, but only two studies were identified, both of them including adults only (18). Bergsma and Carney (19) conducted a systematic review on health-promoting media literacy education, including interventions focusing the analysis and evaluation of health messages in the media. The review found 28 interventions administered to children, adolescents and college students. Due to the differences between studies regarding health issues, research design, theoretical models and outcome measures no overall intervention effectiveness could be calculated. A narrative summary indicated that at least some studies found positive effects of education on participants' critical media skills although none of the studies targeted scientific evaluation of messages.

The previous research indicate that interventions targeting health-related literacy interventions may have positive effects on the students, but no reviews have as of yet looked at enhancing the adolescents abilities in critical appraisal of health claims in the media and elsewhere in society.

Objective

The objective of this systematic review was to identify, appraise and synthesize studies examining the effectiveness of educational interventions in schools that aim to enhance adolescents' abilities in critically appraising health claims.

Methods

The review protocol was registered in PROSPERO International prospective register of systematic reviews with Identification number: CRD42015017936. Additionally, the review adheres to recommendations of the Cochrane Collaboration (20) and the PRISMA checklist for reporting systematic reviews (21) (See Supplementary material S2).

Eligibility criteria

Participants

We included studies on children and adolescents aged 11 to 18, which usually corresponds to grades 6 to 12 in middle school, secondary school, high school or other equivalent educational institutions. Studies that included 10-year olds at sixth grade or 19-year olds at 12th grade were included. On the other hand, studies of the included age group in undergraduate education (e.g. college, university) were excluded.

Studies aimed at teachers were only included if relevant student outcomes were measured.

Interventions

We included any type of school-based educational intervention which aimed to improve students' ability to critically appraise health claims and information through advancing their knowledge about science. We used Ryder's framework (22) as a basis for the teaching and learning of knowledge about science in compulsory school science, and considered interventions that included training on one or more of the following areas:

- Study design (e.g. experimental studies, blinding, placebos, control groups, observational studies)
- Assessing the quality of data (e.g. measurement variability)
- Interpretation of data (e.g. distinction of correlation and causation, sample size and sampling errors)
- Uncertainty in science (e.g. complexity of variables, restrictions on study designs, estimates of risks)
- Science communication (e.g. the role of peer review, funding issues, deficiencies in media reports of research findings)

The educational intervention had to involve claims and information about the human body and health, including conventional medical treatments, complementary and alternative treatments, health conditions, diseases, and physical or mental well-being. We also included studies where health topics were only a means to an end, that is, studies that primarily aimed to enhance students' knowledge about science, and health-related cases served as examples in lessons and assessment.

No restrictions were made with respect to teaching and learning method, educational content and materials, intervention dosage, or who administered the intervention. Furthermore, we included studies where the educational intervention was part of a complex intervention or larger study, and it was possible to separately extract results from that specific intervention.

We excluded studies on regular health education interventions, such as teaching about the benefits of healthy eating or the dangers of smoking. Similarly, we excluded studies of interventions aimed at increasing students' subject matter knowledge (22), for example basic principles of gene inheritance or the workings of the human organ system. Studies on health-related media literacy interventions, involving critical examination of media messages, were excluded unless they contained teaching elements related to knowledge about science as defined above.

Comparisons

The comparison group could receive regular classes, no intervention or another type of intervention, for example a generic health education intervention with no focus on critical appraisal, or studies that compared different methods of delivery, educational contents, intervention dosages or the like.

Outcomes

Primary outcomes included critical appraisal abilities within one or more of the following domains (23):

1. *Knowledge and understanding*: students' retention of facts and concepts related to critical appraisal, for example recognising the need for control groups to justify a health claim about causality; or understanding that a health claim can never be proven, and accordingly health decisions may need to be based on estimates of risk.

2. *Skills*: ability to apply knowledge, for example being able to identify deficiencies in a media report about a health risk.

3. *Behaviour*: transferring the knowledge and skills specified above to everyday situations, for example when sifting through web pages for information on a health problem or lifestyle issue.

Secondary outcomes included:

1. *Attitudes*, students' values/beliefs related to the importance and usefulness of critical appraisal to inform decisions about health.

2. *Participation or completion*, attendance at and reactions on the learning experience, for example participation in class, time spent on class activities, and satisfaction with the educational intervention.

For the primary outcome domains we included self-reported as well as objective outcome measures. Furthermore, studies using both validated and non-validated measurement instruments were included. We assessed potential bias due to use of non-validated instruments during risk of bias assessment.

Types of studies

We included randomised and non-randomised controlled trials that allocated students individually or in clusters (i.e. teachers, classrooms, schools), with pre-and post-test or post-test only, and interrupted time series.

Search strategy

The following databases were searched from their inception through February 2014: MEDLINE, Embase, PsycINFO, AMED (via Ovid), Cinahl, Teachers Reference Centre, LISTA (via EBSCOhost), ERIC, Sociological Abstracts, Social Science Abstracts (via ProQuest), The Cochrane Library (via Wiley), Science Citation Index Expanded and Social Sciences Citation Index (via Web of Science).

To identify grey literature we searched OpenGrey, Social Care Online, Social Science Research Network Library and Google Scholar through June 2014. We searched Clinicaltrials.gov and the International Clinical Trials Registry Platform Search Portal through August 2014 for ongoing studies. Additionally, we searched reference lists of identified relevant reviews and a citation search on included studies to identify additional potentially relevant references.

MWG and LVN developed a highly sensitive search strategy for MEDLINE and ERIC by combining index terms and text words relevant to the population and intervention. MWG modified the search strategy for the other databases and run all searches. A search filter was

applied where appropriate. No language restrictions were applied, although translation of studies in languages other than English, Norwegian, Swedish, Danish and German was beyond the scope of this systematic review (See Supplementary material S1).

Study selection

One reviewer (MWG) performed an initial screening of references identified by the search strategy excluding obviously irrelevant studies. The remaining references were screened independently by two reviewers (MWG and LVN). The same reviewers independently screened potentially relevant references in full text. Any disagreements were resolved by consensus or by involving a third reviewer.

Data extraction

Two reviewers (MWG and LVN) independently extracted data from included studies using a standardised data extraction form. The following data were extracted: methods, setting, student and education provider characteristics, interventions and comparisons (e.g. learning objectives, teaching contents, frequency), outcomes and results. Any disagreements were resolved by consensus. If necessary, we contacted study authors for additional information. Authors' responses varied in details, mainly because of the time passed since studies were executed.

Assessment of risk of bias in included studies

Two reviewers (MWG and LVN) independently assessed risk of bias in included studies by using a modified version of the Cochrane risk of bias tool. Modifications were based on guidelines of the Cochrane Consumers and Communication Review Group (24) and the ACROBAT guidelines for non-randomised studies (25). We assessed risk of bias on ten domains: Sequence generation, allocation concealment, comparability of baseline characteristics and outcome measurements, blinding of students and education providers, blinding of outcome assessment, departures from intended interventions, incomplete outcome data, selective outcome reporting, outcome measures reliable, and other sources of bias. Each domain was assessed as low; unclear; or high risk of bias. Disagreements were solved by consensus or by involving a third reviewer.

Synthesis of results

We attempted to meta-analyse the study results but due to differences in interventions, designs and insufficient reporting of study results this was considered inappropriate. Thus, we synthesized results descriptively. RevMan 5.3 vas used to recalculate effect estimates if this improved their reporting. We used GRADE (Grading of Recommendations, Assessment, Development and Evaluation) (26) to assess and grade the overall quality of evidence for each outcome).

Results

The literature search identified a total of 17,362 unique references. Due to the sensitivity of the search a great number of these references were obviously irrelevant and could be excluded by title only. One reviewer (MWG) preformed this initial screening, removing 9223 references, eliminating roughly 55% of the identified references. In case of any doubt, references were not excluded at this stage.

Two reviewers (LVN and MWG) independently screened the title and/or abstract of the remaining 8139 references. 269 references were considered potentially relevant and obtained in full text. We excluded 263 publications. Reasons for exclusion are provided for the publications that most likely would have been expected to be among the included studies, as recommended by EPOC (27). (See Supplementary material S3). The selection process is outlined in Figure 1.

Study characteristics

Five studies from six publications were included in the review (28-33). The studies reported different summary statistics, and only two studies reported their results in adequate detail. This limitation, in addition to the substantial variations in interventions, made it unfeasible to conduct meta-analyses. Interventions across studies could be classified into two main comparisons: *Educational interventions comparing different teaching modalities* and *educational interventions compared to instruction as usual*. Table 1 and 2 give descriptions of included studies within comparisons and summary of findings are provided in (See Supplementary material S4)

Setting

Four of the studies took place in lower and upper secondary schools in the US (28-32), the fifth study in upper secondary schools in Germany (33).

Student participants

The total number of students across four of the studies is 908 (28-30, 32, 33). One study (31) only provided the number of participating classes (n=9).

All studies included both female and male students, grade levels ranged from 7th to 12th grade. Student populations in the four US studies were ethnically diverse (28-32), the majority of students came from low- or middle income households (28-30, 32). In the German study the mean average with migration background was 16%, socioeconomic status was not reported (33). Students' school performance was mid-range (32) or low (28, 29, 32, 33), in one study the intervention group was students with learning disabilities whose achievement levels ranged 2nd to 10th grade (32).

Content and delivery of interventions

The interventions addressed miscellaneous health topics and varied substantially in terms of scientific topics covered. Using Ryder's framework for knowledge about science (22), we nonetheless found some similarities across studies. All studies addressed aspects of study designs and data interpretation, the use of control variables and differences between causality and correlation being common topics across studies. Four studies addressed science communication (30-33), most often related to deficiencies in media reports of science.

The pedagogical principles underpinning curriculum development and teaching methods varied across studies. Irrespective of pedagogical perspective, the use of active learning rather than traditional lecturing was a central tenet of interventions in all studies. Active approaches took various forms such as small-group work and investigations (28-31, 33), or worksheets (30, 33). Another predominant feature was authentic problem-solving to engage students in the learning process.

There was generally little information about the education providers in the studies in terms of age, years of experience, and competence in the area studied. In two of the studies the researchers themselves delivered the whole of substantial parts of the intervention (31, 33), in another two studies the teachers received training before delivering the intervention studied.

Reported outcomes

Three studies assessed knowledge and skills relevant for critical appraisal, for instance understanding of epidemiological research (28-30, 33). Three of the studies assessed critical appraisal-related outcomes more directly in terms of applying causal reasoning to constructed health scenarios or authentic news reports of research (31, 32). All studies measured outcomes immediately or shortly following the intervention, and only two studies used pre- and post-intervention assessment of outcomes (30, 31). Overall, outcomes were measured using instruments developed in-house by the researches themselves and were considered having reasonable reliability measures. However, information about validity was sparse. None of the studies assessed behaviour, attitudes, or satisfaction in any manner relevant to the review.

Risk of bias assessment

Using the GRADE criteria, we judged the quality of evidence to be very low for all outcomes. All comparisons were downgraded because of a high or unclear risk of bias (See Figure 2 and Supplementary material S5). Additionally, indirectness and imprecision was a problem in most studies (See Supplementary material S4). This was due to studies being mostly undertaken in one single school and because outcomes were mostly addressed in one study only.

Effects of educational interventions

Educational interventions comparing different teaching modalities

We identified only one study, a randomized controlled trial of individual students in one school (28, 29), that compared different teaching modalities. In this study, a four-lesson situated instruction model was compared to an abstracted instruction model for teaching seventh grade students how to determine causality in research studies. Students' causal understanding when applied to constructed reports of health research was higher in the situated learning than in the abstracted learning group (mean difference in percentage score was 14.68%, 95% CI: 9.68 to 19.68, $p < 0.01$).

A number of students in both groups received an additional lesson about how to transfer their causal understanding when making judgments about authentic research reports in contemporary media. Two weeks after instruction, only two students in the situated group were able to transfer their learning. It should be noted that the original randomization was broken as only a selection of class periods that met for either the situated or abstracted instruction were taught for transfer (See Table 1) (28, 29).

The quality of evidence for the results for this comparison was graded very low (See Supplementary material S4).

Educational interventions compared to instruction as usual

Four studies compared various educational interventions to instruction as usual. One study was a cluster-randomised controlled study with teachers as the allocation unit (30). The other three studies were non-randomised controlled studies with teachers (31, 32) or classes (33) as the unit of allocation. All interventions comprised at least 15 lessons of instruction and spanned three weeks to six months (Table 2).

Kaelin and colleagues (30) tested the effectiveness of an epidemiology curriculum comprising 34 lessons for seventh-grade students. Epidemiological understanding and skills were evaluated using self-reports (questionnaire) and a direct testing (multiple-choice test). Study authors provided results for sub-groups based on the experience of the intervention teachers and the number of lessons taught. When objectively assessed, there were small improvement in epidemiological knowledge and skills, but improvements were not statistically significant for students receiving less than 10 lessons ($p > 0.05$). Overall, students mean scores, across all groups were generally low (below 50% correct answers).

Derry and colleagues (31) evaluated an instructional unit in causal reasoning for eight-grade students. A central component was simulation (role play) of a legislative hearing about regulating the dietary supplement industry; Students' skills were assessed using a written test judicial dialogue followed by a question requiring causal reasoning Intervention classrooms gained a higher causal reasoning score compared to the control classrooms (mean difference in adjusted posttest scores was 1.34 points difference reported to be statistically significant, no CIs or p-values provided).

Leshowitz and colleagues (32) compared a causal reasoning instruction for special education (SE) students in grade 7 to 12 to instruction as usual for general education students in grade 11. The intervention and preceding test emphasized the principles of causality as applied to lay information sources, including advertisement claims and news reports of research. Special education students' test scores exceeded scores of the control group of general students (mean difference was 1.26 points, $p < 0.01$).

Steckelberg and colleagues (33) pilot tested an extensive curriculum on aspects of evidence-based medicine for eleventh-grade students. Teaching modules comprised study designs for evaluating effects of interventions and diagnostic testing, critical appraisal of RCTs and systematic reviews. Students' knowledge and skills were assessed using a test that

measured competencies in subareas such as basic statistics and experimental design (34). Competencies in the intervention group, measured as person parameters (Rasch model), were statistically compared to the control group at posttest (mean difference in person parameters was 114, 95% CI: 86 to 142, $p < 0.01$). A difference in 100 person parameters was considered relevant (33).

The quality of evidence for all results within this comparison was graded very low (See Supplementary material S4).

Discussion

Our aim was to systematically review the effects of school-based educational interventions for enhancing adolescents' abilities in critical appraisal of health claims. Despite an extensive literature search, only five studies from six publications met the inclusion criteria. The studies evaluated interventions that varied considerably in their scope, topics coverage, delivery and dosage. Furthermore, they measured knowledge and skills relevant for critical appraisal or critical appraisal-related outcomes more directly. Overall, using GRADE, the quality of evidence for all outcomes was very low. Thus, our confidence in the results is weak. Nevertheless, to our knowledge this is the first systematic and transparent approach to the topic and thus we believe that the review makes an important contribution to the field.

One of the main reasons for downgrading was the predominately high risk of bias in studies, which means there is a risk that they overestimate or underestimate the true intervention effect. Sequence generation and allocation concealment were generally a problem across studies due to non-random allocation or insufficient reporting procedures used. This makes comparability of groups in the studies questionable. Still, all studies shoved promising effects in favour of the interventions being tested. Although effects were reported to be statistically significant, the differences between groups in two of the studies were small and their scores considerably below the medium level score no matter the instruction received (30, 31). Moreover, a statistically significant result does not imply it being educationally significant. Only one study reported what would be considered a relevant educational change in posttest scores (33). Nevertheless, Hedges and Hedberg (35) state that effect estimates as small 0.20 can be of interest in an educational context. Analyses of randomised studies in education suggest that statistically small sizes are quite common, with average sizes for interventions targeting entire classroom as small as 0.18 and somewhat larger (0.40) in one-to-one interventions (36).

For an educational intervention to have the desired effect, the gain in skill, knowledge or behaviour would have to be retained not just in the time following directly after the intervention (37). None of the included studies evaluated the long-term effects of the interventions. If gains are lost only weeks after the instruction an educational intervention will likely not be deemed successful, even if the immediate results are positive. In her randomised controlled study, Hendricks (28, 29) found a statistically significant difference in mean causal understanding in favour of the situated learning group, the effect size being quite large (0.7). However, two weeks later, only two of the students were able to transfer their learning when faced with an authentic health claim. This probably reflects the rather advanced level of skills being tested and also the short-term status of the intervention. To sustain learning effects students need to practice skills over time. A further note should be made to the point that none of the studies measured students' appraisal behaviour in everyday contexts outside the classroom. This is perhaps not surprising given that most school-based educational interventions or studies of students' performance are mostly concerned with measuring cognitive learning outcomes (15).

Two of the studies were randomised controlled studies, in which individual students (28, 29) or teachers (30) were allocated to groups respectively. Both studies included seventh-grade grade students in the US, but are otherwise good examples of the diversity in settings, interventions and conditions for implementation across studies. Hendricks compared a short-term four-lesson situated instruction to abstracted instruction in causal reasoning in one single school. The two modalities of instruction were implemented in a rather controlled environment to ensure fidelity of implementation (28, 29). The study by Kaelin and colleagues included eight teachers in eight schools, and their respective students. It compared a long-term 34-lesson epidemiology curriculum to usual instruction. Implementation varied considerably among teachers, and none of them taught more than approximately half the lessons at most, despite received training in delivery of lessons beforehand. This reflects challenges in implementing new curricula as well as testing their effectiveness in real classroom situations, not at least due to time pressures on teachers (30).

Despite the differences in contents, teaching methods and intensity of the interventions, there were some overall similarities in the topics covered, the differences between causality and correlation and use of control variables being an area of learning across studies. Moreover, the use of active learning approaches such as small group work and contexts resembling authentic situations were central tenets in the interventions provided. Studies of such approaches in science classrooms show mixed results. In a systematic review,

Bennet, Luben and Hogarth (38) examined the effect of context-based teaching compared to conventional approaches on students' understanding and attitudes to science, and found positive effects on attitudes only. The same authors also reviewed the evidence with regard to small group discussions, demonstrating a positive effect on students' scientific understanding (39). However, the authors urged cautious interpretation of the findings since the researchers conducting the studies were advocates of the method.

No matter the content and methods used, the influence of the teacher on students' learning should not be underestimated. The included studies generally lacked information regarding teacher characteristics and teaching styles, with the study by Hendricks (28, 29) being an exception. John Hattie, citing a study by Slater, Davies and Burgess (40), states that a student in the class of a high-impact teacher has almost one year's advantage over a student in a lower-effects teacher's class (41, p. 23). Although no easy answer exists regarding what constitutes a high-impact teacher, it is within good reason to expect that the teacher has some influence on the effects of the instruction no matter the content.

While the teachers most likely affect the delivery of the educational intervention, the students most likely affect the reception (42). Information about baseline characteristics in terms of grade level, age, ethnicity, and socioeconomic status was provided in all included studies, but generally not in sufficient detail. This made it difficult to judge comparability of study groups. Although the study authors provided as much information as possible when contacted, much of this information was lost over the years or never collected initially. In two studies data was provided for the whole student population, and not specified for intervention and control group (28, 29, 32). Only two studies pretested students, using scores to adjust the posttest results (30, 31). Using a pretest is strongly recommended in intervention studies because it is by far the only way to determine if changes really occur at posttest (43). Absence of student characteristics as well as a pretest was considered a severe limitation in all studies.

Limitations

A comprehensive and sensitive search strategy was used to increase the chance of finding relevant studies, but there is always a chance that studies have been missed due to limitations in database interfaces, inconsistent indexing and wrong choice of text-words and subject terms. Additionally a hand search of relevant scientific journals could have supplemented the search.

The extensive search generated a vast number of references and because of that only one review author did the preliminary screening. Even though this screening only excluded

the obvious irrelevant references, potential studies could have been missed as a result of screening fatigue.

Due to the low number of included studies we have not estimated the risk of publication bias statistically (20). Nonetheless, publication bias may exist. All of the included studies showed results in favour of the educational intervention, but that does not necessarily entail that any educational intervention will induce greater critical appraisal abilities. For several of the outcomes the results were only just statistically significant. Even though the review authors believe the scarcity of relevant studies is based on a generally unexplored area, it is possible that studies showing no effect, or even a negative effect, have not been published.

Conclusions

Implications for practice

The small number of included studies, the great diversity between the studies as well as severe limitations in methodological quality makes it difficult to draw any definitive conclusions concerning the effect of school-based educational interventions for enhancing adolescents' abilities to critically appraise health claims. Nevertheless, absence of evidence is not evidence of absence. Although this systematic review highlights poor evidence in the area there is no grounds for discontinuing efforts in schools to increase young students' appraisal abilities.

Implications for research

More and better-designed studies are needed to assess the effects of school-based educational interventions for enhancing critical appraisal abilities of health claims among adolescents. Ideally, the effects of such interventions should be assessed in a multicentre randomised controlled trial that adheres to high methodological standards such as allocation concealment and protection of contamination. A wider variety of school-based settings would contribute to broadening our understanding of such interventions. Participants, interventions, outcomes and results should be sufficiently reported to allow replication. Moreover, pre-and post-intervention outcome assessment using comparable, reliable and validated outcome measures is required to permit more firm conclusions regarding the effect. Changes in outcomes that are educationally relevant should be defined in advance.

Considering both possible primary gains in students' knowledge or skills and secondary health-related gains, effective school-based interventions aimed at enhancing critical appraisal skills when faced with health claims in society has the potential of reaching and affecting a great number of people

Conflict of interest

The authors have no conflicts of interest to declare.

Acknowledgements

The authors would like to thank Claire Stansfield from The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) for her valuable help in developing the search strategy by reviewing the original Medline strategy. Further we would like to thank the study authors who kindly provided additional information on their studies.

References

1. Kaiser Family Foundation. Generation M2: Media in the Lives of 8- to 18-Year-Olds. Menlo Park, California: 2010.
2. Borzekowski DL, Rickert VI. Adolescent cybersurfing for health information: a new resource that crosses barriers. Arch Pediatr Adolesc Med. 2001;155(7):813-7.
3. Jimenez-Pernett J, de Labry-Lima AO, Bermudez-Tamayo C, Garcia-Gutierrez JF, del Carmen Salcedo-Sanchez M. Use of the internet as a source of health information by Spanish adolescents. BMC Med Inform Decis Mak. 2010;10:6.
4. Skinner H, Biscope S, Poland B, Goldberg E. How Adolescents Use Technology for Health Information: Implications for Health Professionals from Focus Group Studies. J Med Internet Res. 2003;5(4):e32.
5. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med. 2005;60(7):1467-78.
6. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. The Internet: a window on adolescent health literacy. J Adolesc Health. 2005;37(3):243.
7. Tsai P-Y, Chang W-H, Chen S, Chang H-P. Young Adolescents' Intentional Use of Science News. Int J Sci Educ B. 2013:1-24.
8. Manganello JA. Health literacy and adolescents: a framework and agenda for future research. Health Educ Res. 2008;23(5):840-7.
9. Paakkari L, Paakkari O. Health literacy as a learning outcome in schools. Health Educ. 2012;112(2):133-52.
10. Sorensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012;12:80.
11. Horsley T, Hyde C, Santesso N, Parkes J, Milne R, Stewart R. Teaching critical appraisal skills in healthcare settings. Cochrane Database Syst Rev. 2011;(11):CD001270.
12. Pettersen S. Critical thinking in Norwegian upper secondary biology education: The cases of complementary-alternative-medicine and health claims in the media. NorDiNa. 2005;1(2):61-71.
13. Cooper BEJ, Lee WE, Goldacre BM, Sanders TAB. The quality of the evidence for dietary advice given in UK national newspapers. Public Underst Sci. 2012;21(6):664-73.

14. Glenton C, Paulsen EJ, Oxman AD. Portals to Wonderland: health portals lead to confusing information about the effects of health care. BMC Med Inform Decis Mak. 2005;5:7.
15. OECD. PISA 2015. Draft Science Framework. Paris: 2013.
16. Ryder J. Identifying Science Understanding for Functional Scientific Literacy. Stud Sci Educ. 2001;36(1):1-44.
17. Millar R. Twenty First Century Science: Insights from the Design and Implementation of a Scientific Literacy Approach in School Science. Int J Sci Educ. 2006;28(13):1499-521.
18. Car J, Lang B, Colledge A, Ung C, Majeed A. Interventions for enhancing consumers' online health literacy. The Cochrane database of systematic reviews. 2011;(6):CD007092.
19. Bergsma LJ, Carney ME. Effectiveness of health-promoting media literacy education: a systematic review. Health Educ Res. 2008;23(3):522-42.
20. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updates March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org
21. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
22. Ryder J. School science education for citizenship: Strategies for teaching about the epistemology of science. J Curriculum Stud. 2002;34(6):637-58.
23. Kirkpatrick DL. Evaluation of training. In: Craig L, Bittel LR, editors. Training and Development Handbook. New York: McGraw-Hill; 1967. p. 87-112.
24. Ryan R, Hill S, Prictor M, McKenzie J. Cochrane Consumers and Communication Review Group. Study Quality Guide. May 2013http://cccrg.cochrane.org/authorresources [accessed May 14th 2015]
25. Sterne JAC, Higgins JPT, Reeves BC, on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT- NRSI), Version 1.0.0, September 2014. Available from: http://www.riskofbias.info. [accessed May 5th 2015]
26. McMaster University, Evidence Prime Inc. GRADEpro Guideline Development Tool. 2015.
27. Effective Practice and Organisation of Care (EPOC). Which studies should be included in the Characteristics of excluded studies table? EPOC Resources for review authors Oslo: Norwegian Knowledge Centre for the Health Services; 2015. Available from:

http://epoc.cochrane.org/epoc-specific-resources-review-authors [accessed May 4th 2015]

28. Hendricks CC. Teaching causal reasoning through cognitive apprenticeship: What are results from situated learning? J Educ Res. 2001;94(5):302-11.
29. Hill CC. The effects of situated learning, abstracted instruction, and teaching for transfer on students' use of statistical reasoning to solve real-world problems. Diss Abstr Int Section A: Humanities and Social Sciences. 1998;59(5-A):1457.
30. Kaelin MA, Huebner WW, Nicolich MJ, Kimbrough ML. Field Test of an Epidemiology Curriculum for Middle School Students. Am J Health Educ 2007;38(1):16-31.
31. Derry SJ, Levin JR, Osana HP, Jones MS. Developing middle-school students' statistical reasoning abilities through simulation gaming. In: Lajoie SP (Ed). Reflections on statistics: Learning, teaching, and assessment in Grades K-12. Mahwah: Lawrence Erlbaum Associates Publishers; 1998. pp. 175-95.
32. Leshowitz B, Jenkens K, Heaton S, Bough TL. Fostering critical thinking skills in students with learning disabilities: an instructional program. J Learn Disabil. 1993;26(7):483-90.
33. Steckelberg A, Hülfenhaus C, Kasper J, Mühlhauser I. Ebm@school--a curriculum of critical health literacy for secondary school students: results of a pilot study. Int J Public Health. 2009; 54(3):158-65
34. Steckelberg A, Hulfenhaus C, Kasper J, Rost J, Muhlhauser I. How to measure critical health competences: development and validation of the Critical Health Competence Test (CHC Test). Adv Health Sci Educ Theory Pract. 2009;14(1):11-22.
35. Hedges LV, Hedberg EC. Intraclass Correlation Values for Planning Group-Randomized Trials in Education. Educational Evaluation and Policy Analysis. 2007;29(1):60-87.
36. Lipsey MW, Puzio K, Yun C, Hebert MA, Steinka-Fry K, Cole MW, et al. Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education; 2012.
37. Andrews R. The Place of Systematic Reviews in Education Research. Brit J Educ Stud. 2005;53(4):399-416.
38. Bennett J, Lubben F, Hogarth S. Bringing Science to Life: A Synthesis of the Research Evidence on the Effects of Context-Based and STS Approaches to Science Teaching. Sci Educ. 2007;91(3):347-70.

39. Bennett J, Hogarth S, Lubben F, Campbell B, Robinson A. Talking Science: The research evidence on the use of small group discussions in science teaching. Int J Sci Educ. 2010;32(1):69-95.
40. Slater H, Davies NM, Burgess S. Do Teachers Matter? Measuring the Variation in Teacher Effectiveness in England. Oxford B Econ Stat. 2012;74(5):629-45.
41. Hattie J. Visible learning for teachers : maximizing impact on learning. Hoboken, N.J: Taylor & Francis; 2011.
42. Patricio M, vaz Carneiro A. Systematic reviews of evidence in medical education and clinical medicine: is the nature of evidence similar? Med Teach. 2012;34(6):474-82.
43. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.

Figure 1

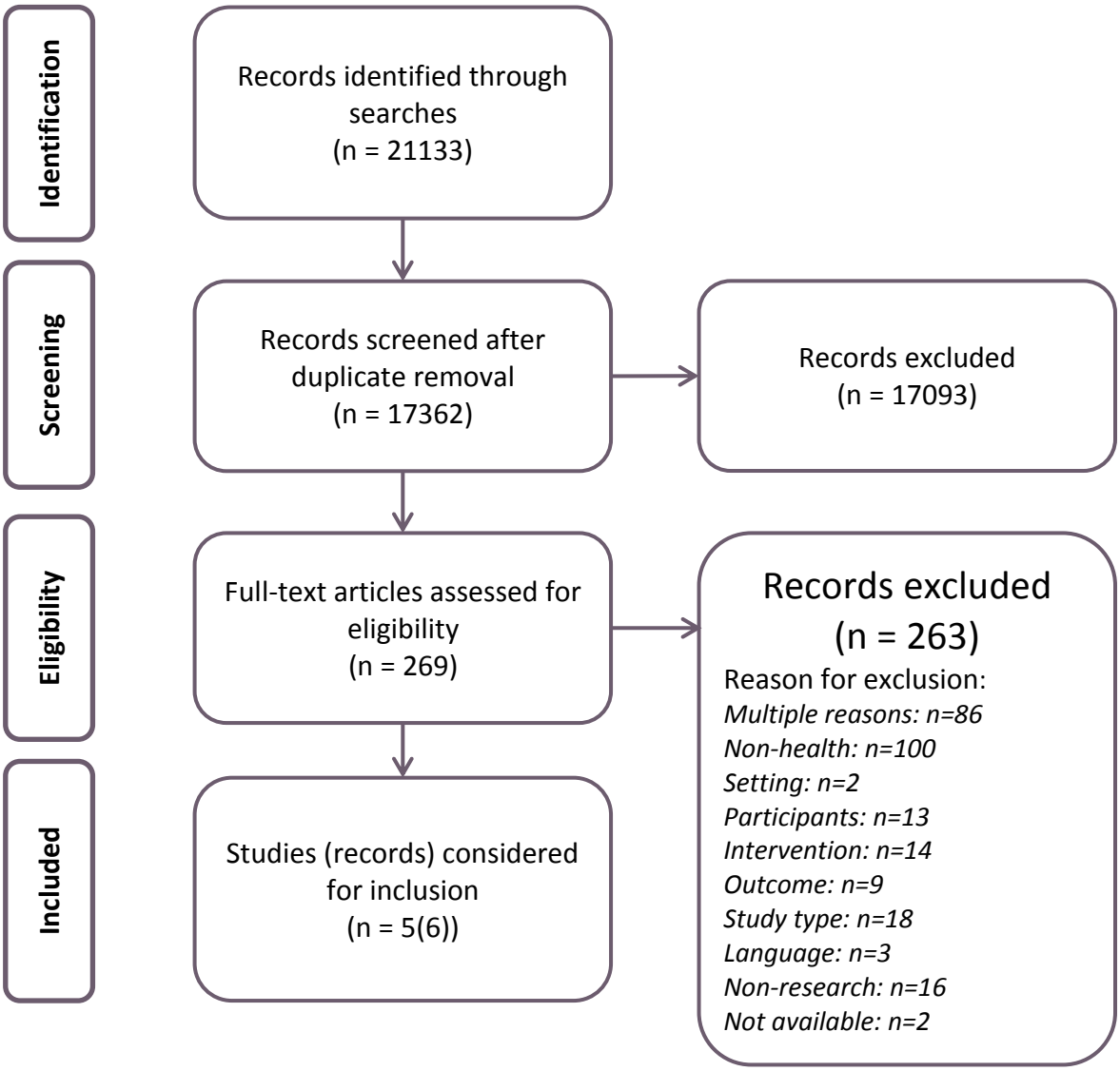
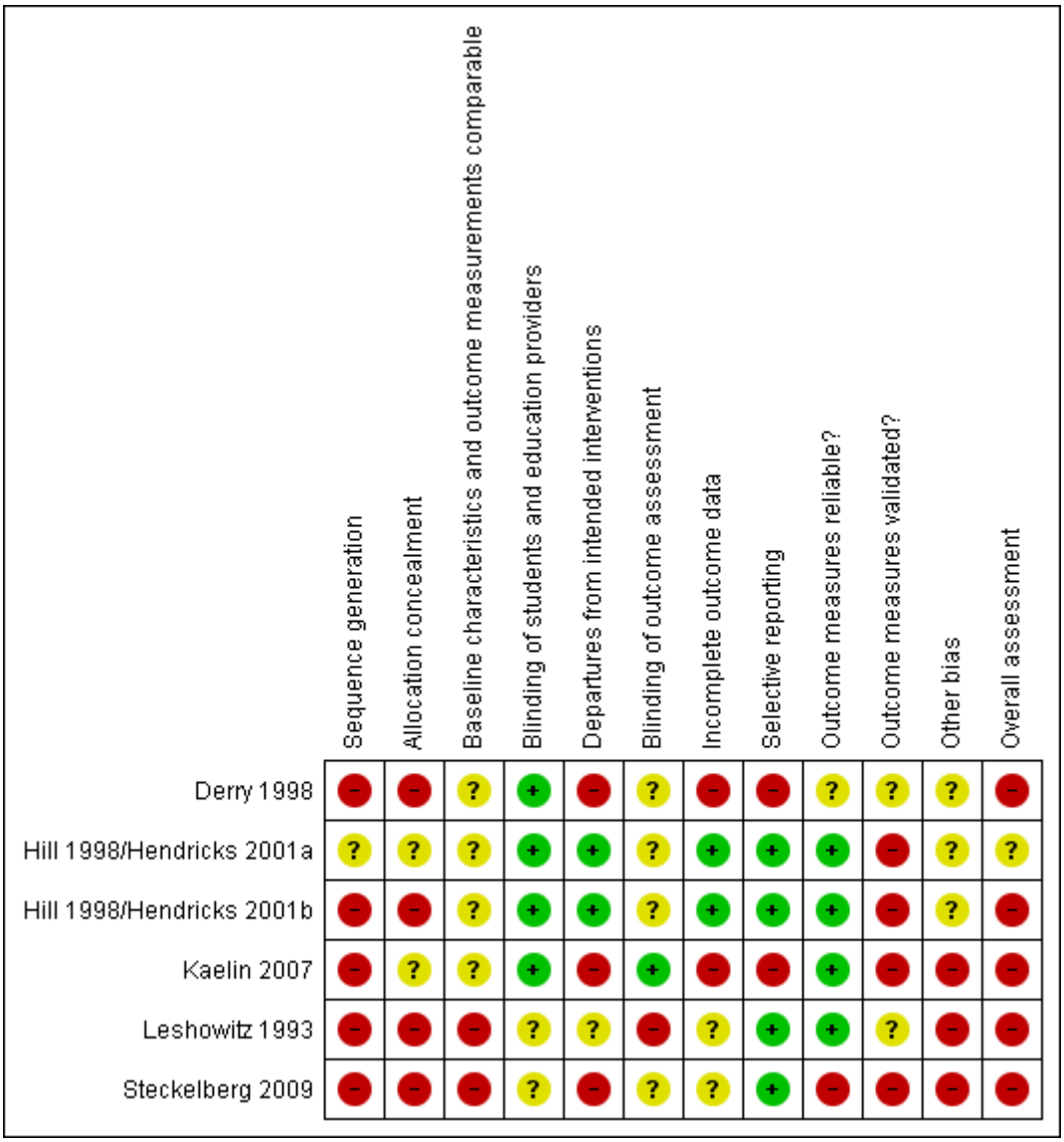


Figure 1 PRISMA Flow chart of search results and screening process

Figure 2



a: situated versus abstracted instruction
b: transfer instruction

Figure 2 Risk of bias summary for included studies

Table 1 School-based educational intervention: Comparing different teaching modalities for enhancing adolescents' critical appraisal abilities

Study, *design*, allocation unit [Ref]	**Setting**	**Students**	**Education providers**	**Science area (Ryder (27))**	**Intervention topics**	**Dosage**	**Pedagogical framework, teaching methods**	**Outcome [domain], type of measurement**
Hill 1998 / Hendricks 2001 (33, 34) *Randomised controlled study with posttest only*; Students	One lower secondary school, US	220 students in 7th grade Males (%): 48 Age: Not reported 56% African, 40% Caucasian SES*: Diverse SP**: Low	Two first-year female science teachers (one per instruction group).	Study design Interpreting data Science communication	*Situated instruction in causal reasoning*. Causality vs correlation, role of random assignment, causality cues in reports of research. *Comparison:* Abstracted instruction. Same topics as situated instruction. Health topics (examples): Exercise, stress.	4 lessons á 50 min over 2 weeks for both teaching modalities.	Situated cognition (e.g. Vygotsky). Teacher as guide/mentor rather than lecturer. Classroom or small group discussions, reflective activities using authentic examples. Abstracted instruction: Knowledge transmission, lectures. No/few reflective activities. Constructed examples.	Understanding causality [Knowledge | Skills] Test scenarios based on authentic media reports of research. Students were asked to determine cause-and effect variables and use of random assignment. Short-open and selected-response test: 0 to 15 points
Transfer instruction: *Non-randomised controlled study with posttest only*; Class periods		194 of 220 participating students			*Transfer instruction*: Applying causal reasoning in authentic situations (e.g. news reports of research, medical studies).	One lesson á 50 min one week after situated/ abstracted instruction	Small groups, authentic examples.	Causal reasoning [Skills] Authentic news report from CNN ("Grapes inhibit cancer"). Students were asked to judge believability of story and support their conclusions. Open-response test: 0 to 3 points

*SES = Socioeconomic status. **SP = School performance

Table 2 School-based educational interventions compared to instruction as usual for enhancing adolescents' critical appraisal abilities

Study, *design*, allocation unit [Ref]	Setting	Students	Education providers	Science area (Ryder (27))	Intervention topics	Dosage	Pedagogical framework, teaching methods	Outcome, *measure*
Derry 1998 *Non-randomised group study with pre- and posttest*; Teachers (36)	8 classes in one lower secondary school, US	8th grade (no. of students not reported) Males (%): Not reported Age: Not reported Ethnicity: Diverse SES*: Diverse SP**: Diverse	One science and one social studies teacher; expert scientist; project researchers	Study design Interpreting data Uncertainty in science Science communication	*Simulation gaming (role play of legislation hearing) in causal reasoning.* Topics: Single-case observation vs RCTs, governmental regulations of scientific and lay community, valid statistical inference. Health topics (examples): Cancer; dietary supplements; violence	15 lessons á 70 min over 3 weeks	Situated cognition (e.g. Vygotsky), radical constructionism Small groups, lectures + class discussions, extensive project work, teachers as mentors and models	Causal reasoning [Skills] Test scenarios of a court of law with a debate presenting various forms of evidence and counterarguments. Students were asked to answer a question which required causal reasoning. Open-response test: -1 to 13 points
Kaelin 2007 *Non-randomised group study with pre- and posttest*; Teachers (35)	8 lower secondary schools, US	512 students in 7th grade Males (%): 47 Age: Mean 12.3 Ethnicity: 54% Hispanic, 37% African American SES: Low SP: Diverse	Seven female and two male science teachers	Study design Interpreting data Uncertainty in science	*Epidemiology curriculum.* Topics: Descriptive and analytical epidemiology, flaws in observational studies, societal role of epidemiology, evaluating prevention strategies. Health topics (examples): Acne, back pain	34 lessons á 1-2 class periods over six months. No. of lessons taught varied (see Table S2)	Understanding by Design: enduring understandings that have lasting values outside classrooms Small groups (Epi teams investigations), lectures, worksheets, portfolios, pre-/post assessments	Understanding epidemiology [Knowledge | Skills] Self-reported understanding: Likert scale; 5 to 25 points Multiple-choice test: 0 to 11 points
Leshowitz 1993 *Non-randomised group study with posttest only*; Teachers (37)	5 classes in one lower and one upper secondary school, US	55 special (SE) and general education (GE) students in grade 7-12 Males (%): 68 (SE only). Age: 16 – 20 Ethnicity: 68% Caucasian, 32% Hispanic (SE only, GE "closely match") SES: Low SP: Equiv. grade 2-10 for SE, GE not reported.	Two pre-service special education teachers age approx. 30 or 40 respectively	Study design Interpreting data Uncertainty in science	*Causal reasoning instruction.* Topics: Applying principles of causality to advertisements and news reports of research. Independent and dependent variables, control groups, confounding. Health topics (examples): Cancer; aspirin	25 lessons á 45 min over 4-6 weeks	Socratic dialogue format Class discussions using authentic examples.	Causal reasoning [Skills] Constructed test including an advertisement and short news report of a research study. Students were asked to identify claim, graph cause-effect variables, and explain whether data proved the claim. Short-open response test: 0 to 6 points
Steckelberg 2009 *Non-randomised group study with posttest only*; Classes (38)	12 classes in upper secondary schools, Germany	255 students in 11th grade Males (%): 38 Age: Mean 17.5 Ethnicity: 16% non-ethnically German SES and SP not reported	Two project researchers; class teachers offered to be present	- Study design - Assessing data quality - Interpreting data - Uncertainty in science - Science communication	*Evidence-based medicine curriculum.* Topics: Expert vs EBM information, study designs (e.g. RCTs, diagnostic studies), epidemiological statistics, systematic reviews, question formulation and Internet/database searching. Health topics (examples): Nutrition, smoking	22 lessons over one week	Klafki's framework for reflection of aims and instruction Small groups, lectures + class discussions, worksheets, extensive project work using authentic examples	Understanding EBM aspects Critical Health Competence test: Multiple-choice and short-open responses Total score in person parameters (Rasch model)

*SES = Socioeconomic status. **SP = School performance.

Supplementary material 1

Search history studies on the effects of educational interventions to enhance critical health literacy in adolescence

Developed by: Lena V. Nordheim and Malene W. Gundersen
Executed by: Malene W. Gundersen

Total number of results with dublicates: 19717
Total number of results without dublicates: 15946

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
Date: 25.02.2014
Hits: 2668

#	Searches	Results
1	(age? adj2 ("11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)).tw.	182796
2	Schools/	20033
3	School Health Services/	13326
4	Students/	33119
5	(schoolchild* or school-child* or school student* or pupil?).tw.	51078
6	((middle or secondary or high) adj school?).tw.	27364
7	((six or sixth or seven* or eight? or nine or ninth or ten or tenth or eleven* or twelve or twelfth) adj3 grade*).tw.	8159
8	(grade? 6 or grade? 7 or grade? 8 or grade? 9 or grade? 10 or grade? 11 or grade? 12).tw.	3130
9	(year? 6 or year? 7 or year? 8 or year? 9 or year? 10 or year? 11 or year? 12 or year? 13).tw.	15082
10	or/2-9	138068
11	Education/	18114
12	Health education/	51329
13	Health Education, Dental/	5709
14	Health Fairs/	427
15	Sex Education/	7668

16	Patient Education as Topic/	68798
17	Education, Distance/	2555
18	Education, Nonprofessional/	90
19	Education, Special/	8065
20	Competency-Based Education/	2693
21	Curriculum/	56816
22	exp Programmed Instruction as Topic/	11445
23	Teaching/	41488
24	exp Teaching Materials/	93661
25	exp Educational Technology/	84927
26	ed.fs.	215110
27	(educat* or train* or teach* or workshop? or work-shop? or seminar? or course? or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson? or taught or module?).tw.	1771648
28	or/11-27	2007926
29	Consumer Health Information/	1629
30	Information Services/	14906
31	Research/	173406
32	Empirical Research/	2356
33	Research Design/	76065
34	Research Report/	818
35	Statistics as Topic/	82185
36	Periodicals as Topic/	35836
37	Epidemiology/	11411
38	Pharmacoepidemiology/	1209

39	Pamphlets/	3089
40	Newspapers/	2465
41	Mass Media/	8654
42	Television/	11246
43	Radio/	1995
44	exp Internet/	47870
45	or/29-44	443454
46	Information Seeking Behavior/	589
47	Problem Solving/	20732
48	Problem-Based Learning/	5169
49	Data Interpretation, Statistical/	46821
50	exp Evidence-Based Practice/	58796
51	or/46-50	130683
52	45 and 51	15009
53	Science/ or exp Information Literacy/ or Judgment/ or Decision Making/ or Thinking/	102755
54	52 or 53	116902
55	((health or information or mathematical or quantitative or science or scientific* or media) adj2 (literacy or literate? or illiteracy or illiterate?)).tw.	3181
56	(information adj2 competen*).tw.	176
57	numeracy.tw.	515
58	(scientific adj2 (skill? or think* or reason*)).tw.	677
59	(critical adj2 (think* or reason*)).tw.	2324
60	evidence-based.tw.	54916
61	or/55-60	61274

62	health litera$2.af.	3061
63	medical literacy.af.	23
64	(health and literacy).ti.	1515
65	(functional and health and literacy).tw.	380
66	low-litera$2.ti.	191
67	litera$2.ti.	3784
68	illitera$2.ti.	252
69	reading/	16605
70	comprehension/	7613
71	or/62-70	27696
72	*health promotion/	32764
73	*health education/	29139
74	*patient education/	30424
75	*communication barriers/	2008
76	*communication/	26026
77	*health knowledge,attitudes,practice/	35774
78	*attitude to health/	35657
79	*comprehension/ and *educational status/	35
80	(family and literacy).ti.	33
81	(drug labeling.af. or prescriptions, drug/) and comprehension.af.	129
82	((cancer or diabetes or genetics) and (literacy or comprehension)).ti.	225
83	(limited and (educational status or (educational and status) or literacy)).af.	2723
84	(patient$1 and (educational status or (educational and status) or literacy)).af.	16073
85	(patient$1 and (comprehension or understanding)).ti.	1859

86	or/72-85	193790
87	71 and 86	4672
88	61 or 87	63741
89	((health or consumer or medical or scien*) adj2 information).tw.	26627
90	risk information.tw.	1133
91	((health or medical or scien*) adj2 (claim? or statement? or message?)).tw.	6073
92	((health or medical or scien*) adj2 (paper? or article? or report? or literature? or journal? or periodical? or research)).tw.	83831
93	(media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) adj2 information) or pamphlet? or leaflet? or booklet? or brochure?).tw.	1621304
94	(internet* or world wide web or worldwide web or web site? or website? or web portal? or blog* or web log* or bulletin board? or bulletinboard? or message board? or messageboard? or forum? or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page? or webpage? or chat room? or chatroom?).tw.	57570
95	or/89-94	1774677
96	88 and 95	9375
97	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or consumer or medical or scien*) adj2 information)).tw.	2111
98	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 risk information).tw.	160
99	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or medical or scien*) adj2 (claim? or statement? or message?))).tw.	353
100	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or medical or scien*) adj2 (paper? or article? or report? or literature? or journal? or periodical? or research))).tw.	4333
101	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) adj2 information) or pamphlet? or leaflet? or booklet? or brochure?)).tw.	38291
102	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*)	3916

	adj7 (internet* or world wide web or worldwide web or web site? or website? or web portal? or blog* or web log* or bulletin board? or bulletinboard? or message board? or messageboard? or forum? or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page? or webpage? or chat room? or chatroom?)).tw.	
103	or/54,96-102	171216
104	1 and 28 and 103	741
105	10 and 103	3661
106	104 or 105	4306
107	Randomized Controlled Trial.pt.	363152
108	Controlled Clinical Trial.pt.	87554
109	Multicenter Study.pt.	164632
110	(randomis* or randomiz* or randomly or random allocat*).ti,ab.	538697
111	(group? and (random* or between* or control* or intervent*)).ab.	1444460
112	(multicenter or multi center or multicentre or multi centre).ti. or trial.ti,ab.	366387
113	(intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) adj5 (study or studies or design)) or (before adj5 after) or (pre adj5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect? or effectiveness or impact or time series or time point? or repeated measur*).ti,ab.	7816399
114	Program Evaluation/	44669
115	Pilot Projects/	79858
116	or/107-115	8309951
117	exp Animals/	17053536
118	Humans/	13172587
119	117 not (117 and 118)	3880949
120	116 not 119	6435764
121	Review.pt.	1833187

122	Meta Analysis.pt.	44218
123	News.pt.	159253
124	Comment.pt.	568314
125	Editorial.pt.	346162
126	Letter.pt.	825462
127	cochrane database of systematic reviews.jn.	9846
128	comment on.cm.	568313
129	(systematic review or literature review).ti.	46830
130	or/121-129	3287171
131	120 not 130	5722303
132	106 and 131	2668

Database: Embase 1974 to 2014 February 24 via Ovid
Date: 25.02.2014
Hits: 2832

#	Searches	Results
1	(age? adj2 ("11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)).tw.	257469
2	School/	42332
3	High school/	8661
4	Middle school/	531
5	School health service/	13710
6	student/	54358
7	high school student/	3060
8	middle school student/	429
9	(schoolchild* or school-child* or school student* or pupil?).tw.	63173
10	((middle or secondary or high) adj school?).tw.	34422
11	((six or sixth or seven* or eight? or nine or ninth or ten or tenth or eleven* or twelve or twelfth) adj3 grade*).tw.	10016
12	(grade? 6 or grade? 7 or grade? 8 or grade? 9 or grade? 10 or grade? 11 or grade? 12).tw.	3995
13	(year? 6 or year? 7 or year? 8 or year? 9 or year? 10 or year? 11 or year? 12 or year? 13).tw.	23367
14	or/2-13	203954
15	Education/	307434
16	Course content/	749
17	Curriculum/	62516
18	Curriculum development/	2549
19	Education program/	36454
20	Educational mobility/	89

21	Educational model/	5954
22	Educational technology/	2301
23	exp health education/	238200
24	Interdisciplinary education/	824
25	Learning environment/	1263
26	Problem based learning/	4250
27	Sexual education/	9421
28	exp special education/	1428
29	Student assistance program/	88
30	Study skills/	183
31	Teaching/	65730
32	Health education/	79634
33	Dental health education/	5308
34	nutrition education/	2290
35	School health education/	413
36	Patient education/	88077
37	(educat* or train* or teach* or workshop? or work-shop? or seminar? or course? or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson? or taught or module?).tw.	2267821
38	or/15-37	2540778
39	consumer health information/	2294
40	information service/	14657
41	science/	27435
42	research/	163969
43	empirical research/	3341

44	study design/	13528
45	pharmacoepidemiology/	7145
46	mass medium/	14516
47	television/	13098
48	telecommunication/	17996
49	internet/	75187
50	social media/	1952
51	or/39-50	335870
52	information seeking/	881
53	problem solving/	26171
54	problem based learning/	4250
55	decision making/	140629
56	statistical analysis/	192497
57	evidence based practice/	27395
58	evidence based medicine/	83589
59	or/52-58	463091
60	51 and 59	16210
61	*science/ or health literacy/ or information literacy/ or thinking/ or critical thinking/	32122
62	60 or 61	47438
63	((health or information or mathematical or quantitative or science or scientific* or media) adj2 (literacy or literate? or illiteracy or illiterate?)).tw.	4014
64	(information adj2 competen*).tw.	220
65	numeracy.tw.	668
66	(scientific adj2 (skill? or think* or reason*)).tw.	860

67	(critical adj2 (think* or reason*)).tw.	2648
68	evidence-based.tw.	73444
69	or/63-68	81216
70	health litera$2.af.	4246
71	medical literacy.af.	30
72	(health and literacy).ti.	1803
73	(functional and health and literacy).tw.	527
74	low-litera$2.ti.	232
75	litera$2.ti.	4413
76	illitera$2.ti.	311
77	reading/	32763
78	comprehension/	17495
79	or/70-78	52932
80	*health promotion/	30958
81	*health education/	34748
82	*patient education/	25106
83	*communication disorder/	2781
84	*interpersonal communication/	36626
85	*attitude to health/	41886
86	*comprehension/ and *educational status/	20
87	(family and literacy).ti.	36
88	(drug labeling.af. or prescription/) and comprehension.af.	294
89	((cancer or diabetes or genetics) and (literacy or comprehension)).ti.	305
90	(limited and (educational status or (educational and status) or literacy)).af.	3470

91	(patient$1 and (educational status or (educational and status) or literacy)).af.	21052
92	(patient$1 and (comprehension or understanding)).ti.	2371
93	or/80-92	185680
94	79 and 93	6637
95	69 or 94	84942
96	((health or consumer or medical or scien*) adj2 information).tw.	38601
97	risk information.tw.	1447
98	((health or medical or scien*) adj2 (claim? or statement? or message?)).tw.	8389
99	((health or medical or scien*) adj2 (paper? or article? or report? or literature? or journal? or periodical? or research)).tw.	111750
100	(media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) adj2 information) or pamphlet? or leaflet? or booklet? or brochure?).tw.	2247221
101	(internet* or world wide web or worldwide web or web site? or website? or web portal? or blog* or web log* or bulletin board? or bulletinboard? or message board? or messageboard? or forum? or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page? or webpage? or chat room? or chatroom?).tw.	78679
102	or/96-101	2453226
103	95 and 102	13463
104	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or consumer or medical or scien*) adj2 information)).tw.	2633
105	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 risk information).tw.	201
106	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or medical or scien*) adj2 (claim? or statement? or message?))).tw.	472
107	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or medical or scien*) adj2 (paper? or article? or report? or literature? or journal? or periodical? or research))).tw.	5585
108	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*)	55382

	adj7 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) adj2 information) or pamphlet? or leaflet? or booklet? or brochure?)).tw.	
109	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 (internet* or world wide web or worldwide web or web site? or website? or web portal? or blog* or web log* or bulletin board? or bulletinboard? or message board? or messageboard? or forum? or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page? or webpage? or chat room? or chatroom?)).tw.	5379
110	or/62,103-109	125700
111	1 and 38 and 110	780
112	14 and 110	3745
113	111 or 112	4382
114	randomized controlled trial/	370804
115	controlled clinical trial/	409854
116	multicenter study/	119062
117	pretest posttest control group design/	197
118	pretest posttest design/	637
119	crossover procedure/	40064
120	experimental design/	9139
121	experimental study/	13071
122	pilot study/	74459
123	quasi experimental study/	1778
124	controlled study/	4276516
125	(randomis* or randomiz* or randomly or random allocat*).ti,ab.	729215
126	(group? and (random* or between* or control* or intervent*)).ab.	1974565
127	(multicenter or multi center or multicentre or multi centre).ti. or trial.ti,ab.	507363
128	(intervention* or controlled or control group or compare or comparison* or compared or	9995002

	((prospectiv* or crossover) adj5 (study or studies or design)) or (before adj5 after) or (pre adj5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect? or effectiveness or impact or time series or time point? or repeated measur*).ti,ab.	
129	or/114-128	11883576
130	exp animal/	19845591
131	exp human/	15437873
132	130 not (130 and 131)	4407718
133	129 not 132	9265787
134	"review"/	2096736
135	meta analysis/	81081
136	editorial/	495795
137	letter/	835874
138	("cochrane database of systematic reviews" or "cochrane database of systematic reviews online").jn.	12558
139	(systematic review or literature review).ti.	59885
140	or/134-139	3493690
141	133 not 140	8376743
142	113 and 141	2888
143	remove duplicates from 142	2832

Database: PsycINFO 1806 to February Week 3 2014 via Ovid
Date: 25.02.2014
Hits: 3418

#	Searches	Results
1	(age? adj2 ("11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)).tw.	119546
2	schools/	21356
3	boarding schools/	206
4	charter schools/	322
5	high schools/	4281
6	institutional schools/	357
7	junior high schools/	549
8	middle schools/	2556
9	School Based Intervention/	8580
10	students/	17530
11	high school students/	25871
12	junior high school students/	11229
13	special education students/	6053
14	middle school students/	5756
15	Intermediate School Students/	97
16	(schoolchild* or school-child* or school student* or pupil?).ab,ti.	65086
17	((middle or secondary or high) adj school?).ab,ti.	73952
18	((six or sixth or seven* or eight? or nine or ninth or ten or tenth or eleven* or twelve or twelfth) adj3 grade*).ab,ti.	13653
19	(grade? 6 or grade? 7 or grade? 8 or grade? 9 or grade? 10 or grade? 11 or grade? 12).ab,ti.	6635
20	(year? 6 or year? 7 or year? 8 or year? 9 or year? 10 or year? 11 or year? 12 or year? 13).ab,ti.	2654

21	or/2-20	176380
22	education/	23411
23	distance education/	2383
24	high school education/	2611
25	middle school education/	481
26	private school education/	861
27	public school education/	3839
28	secondary education/	4127
29	special education/	21196
30	curriculum/	18278
31	exp health education/	13924
32	physical education/	3153
33	exp psychology education/	13355
34	science education/	8577
35	social studies education/	1240
36	client education/	2944
37	programmed instruction/	1191
38	exp teaching/	87379
39	curriculum development/	5327
40	educational program planning/	1908
41	exp educational programs/	42482
42	(educat* or train* or teach* or workshop? or work-shop? or seminar? or course? or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson? or taught or module?).ab,ti.	1058174
43	or/22-42	1077954

44	information services/	564
45	sciences/	12182
46	experimental design/	8944
47	Statistical Analysis/	13560
48	Statistics/	6573
49	epidemiology/	37032
50	exp communications media/	46953
51	internet/	20851
52	or/44-51	141166
53	information seeking/	2921
54	decision making/	46423
55	problem solving/	22433
56	problem based learning/	753
57	judgment/	17456
58	thinking/	11899
59	reasoning/	12539
60	evidence based practice/	10107
61	or/53-60	116030
62	52 and 61	4681
63	science achievement/ or literacy/ or health literacy/ or information literacy/ or critical thinking/ or *sciences/ or *reasoning/	30522
64	62 or 63	34343
65	((health or information or mathematical or quantitative or science or scientific* or media) adj2 (literacy or literate? or illiteracy or illiterate?)).ab,ti.	2684
66	(information adj2 competen*).ab,ti.	230

67	numeracy.ab,ti.	811
68	(scientific adj2 (skill? or think* or reason*)).ab,ti.	1223
69	(critical adj2 (think* or reason*)).ab,ti.	4167
70	evidence-based.ab,ti.	21462
71	or/65-70	30128
72	health litera$2.af.	4595
73	medical literacy.af.	26
74	(health and literacy).ti.	712
75	(functional and health and literacy).tw.	243
76	low-litera$2.ti.	138
77	litera$2.ti.	9037
78	illitera$2.ti.	239
79	reading/	13516
80	comprehension/	9868
81	number comprehension/	1056
82	or/72-81	36081
83	*Health Promotion/	11950
84	*health education/	7921
85	*client education/	2427
86	*communication barriers/	220
87	*Communication/	12500
88	*Health Knowledge/	4280
89	*Health Attitudes/	6431
90	*Health Behavior/	12682

91	*academic achievement/	36064
92	*educational attainment level/	2191
93	*comprehension/ and (*academic achievement/ or *educational attainment level/)	76
94	(family and literacy).ti.	196
95	(drug labeling.af. or prescription drugs/) and comprehension.af.	65
96	((cancer or diabetes or genetics) and (literacy or comprehension)).ti.	113
97	(limited and (educational status or (educational and status) or literacy)).af.	11009
98	(patient$1 and (educational status or (educational and status) or literacy)).af.	27463
99	(patient$1 and (comprehension or understanding)).ti.	607
100	or/83-99	121911
101	82 and 100	6572
102	71 or 101	35141
103	((health or consumer or medical or scien*) adj2 information).ab,ti.	7565
104	risk information.ab,ti.	616
105	((health or medical or scien*) adj2 (claim? or statement? or message?)).ab,ti.	2187
106	((health or medical or scien*) adj2 (paper? or article? or report? or literature? or journal? or periodical? or research)).ab,ti.	27906
107	(media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) adj2 information) or pamphlet? or leaflet? or booklet? or brochure?).ab,ti.	219894
108	(internet* or world wide web or worldwide web or web site? or website? or web portal? or blog* or web log* or bulletin board? or bulletinboard? or message board? or messageboard? or forum? or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page? or webpage? or chat room? or chatroom?).ab,ti.	36512
109	or/103-108	281764
110	102 and 109	5189
111	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7	746

	((health or consumer or medical or scien*) adj2 information)).ab,ti.	
112	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 risk information).ab,ti.	95
113	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or medical or scien*) adj2 (claim? or statement? or message?))).ab,ti.	139
114	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or medical or scien*) adj2 (paper? or article? or report? or literature? or journal? or periodical? or research))).ab,ti.	1505
115	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) adj2 information) or pamphlet? or leaflet? or booklet? or brochure?)).ab,ti.	11354
116	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 (internet* or world wide web or worldwide web or web site? or website? or web portal? or blog* or web log* or bulletin board? or bulletinboard? or message board? or messageboard? or forum? or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page? or webpage? or chat room? or chatroom?)).ab,ti.	3139
117	or/64,110-116	53785
118	1 and 43 and 117	678
119	21 and 117	5559
120	118 or 119	6060
121	("0451" or "2000").md.	48686
122	between groups design/	103
123	clinical trials/	7320
124	experimental methods/	8736
125	quasi experimental methods/	112
126	program evaluation/	9866
127	educational program evaluation/	4936
128	treatment effectiveness evaluation/	15883

129	(randomis* or randomiz* or randomly or random allocat*).ti,ab.	90493
130	(group? and (random* or between* or control* or intervent*)).ab.	348167
131	(multicenter or multi center or multicentre or multi centre).ti. or trial.ti,ab.	67861
132	(intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) adj5 (study or studies or design)) or (before adj5 after) or (pre adj5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect? or effectiveness or impact or time series or time point? or repeated measur*).ti,ab.	1594720
133	or/121-132	1726138
134	animal.po.	305361
135	human.po.	2979737
136	134 not (134 and 135)	277988
137	133 not 136	1554756
138	("0800" or "0830" or "1200").md.	107732
139	("Column/Opinion" or "Comment/Reply" or Editorial or "Erratum/Correction" or Letter or Review-Book or Review-Media or Review-Software & Other).dt.	262314
140	(systematic review or literature review).ti.	7903
141	or/138-140	369504
142	137 not 141	1416546
143	120 and 142	3418

Database: AMED (Allied and Complementary Medicine) 1985 to February 2014 via Ovid
Date: 25.02.2014
Hits: 98

#	Searches	Results
1	(age? adj2 ("11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)).ab,ti.	3579
2	schools/	596
3	school health services/	117
4	students/	2240
5	(schoolchild* or school-child* or school student* or pupil?).ab,ti.	683
6	((middle or secondary or high) adj school?).ab,ti.	595
7	((six or sixth or seven* or eight? or nine or ninth or ten or tenth or eleven* or twelve or twelfth) adj3 grade*).ab,ti.	135
8	(grade? 6 or grade? 7 or grade? 8 or grade? 9 or grade? 10 or grade? 11 or grade? 12).ab,ti.	56
9	(year? 6 or year? 7 or year? 8 or year? 9 or year? 10 or year? 11 or year? 12 or year? 13).ab,ti.	455
10	or/2-9	4329
11	education/	3855
12	health education/	411
13	sex education/	56
14	exp patient education/	1659
15	education nonprofessional/	59
16	exp education special/	752
17	curriculum/	1050
18	teaching/	1022
19	(educat* or train* or teach* or workshop? or work-shop? or seminar? or course? or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson? or taught or module?).ab,ti.	42528

20	or/11-19	44569
21	exp information services/	394
22	research/	5728
23	exp models theoretical/	3278
24	research design/	1759
25	exp statistics/	3842
26	epidemiology/	2835
27	exp communications media/	2971
28	internet/	735
29	or/21-28	20219
30	problem solving/	235
31	learning/	1264
32	evidence based medicine/	2256
33	or/30-32	3731
34	29 and 33	596
35	science/ or exp decision making/ or thinking/	3424
36	34 or 35	3995
37	((health or information or mathematical or quantitative or science or scientific* or media) adj2 (literacy or literate? or illiteracy or illiterate?)).ab,ti.	51
38	(information adj2 competen*).ab,ti.	18
39	numeracy.ab,ti.	13
40	(scientific adj2 (skill? or think* or reason*)).ab,ti.	21
41	(critical adj2 (think* or reason*)).ab,ti.	122
42	evidence-based.ab,ti.	2237

43	or/37-42	2445
44	health litera$2.af.	35
45	medical literacy.af.	1
46	(health and literacy).ti.	22
47	(functional and health and literacy).tw.	5
48	low-litera$2.ti.	3
49	litera$2.ti.	130
50	illitera$2.ti.	5
51	reading/	508
52	or/44-51	622
53	Health promotion/	1697
54	Health education/	411
55	Patient education/	1615
56	communication/	2340
57	attitude to health/	2119
58	Educational status/	137
59	(family and literacy).ti.	1
60	((cancer or diabetes or genetics) and (literacy or comprehension)).ti.	1
61	(limited and (educational status or (educational and status) or literacy)).af.	33
62	(patient$1 and (educational status or (educational and status) or literacy)).af.	196
63	(patient$1 and (comprehension or understanding)).ti.	49
64	or/53-63	8204
65	52 and 64	88
66	43 or 65	2500

67	((health or consumer or medical or scien*) adj2 information).ab,ti.	588
68	risk information.ab,ti.	6
69	((health or medical or scien*) adj2 (claim? or statement? or message?)).ab,ti.	139
70	((health or medical or scien*) adj2 (paper? or article? or report? or literature? or journal? or periodical? or research)).ab,ti.	2779
71	(media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) adj2 information) or pamphlet? or leaflet? or booklet? or brochure?).ab,ti.	9891
72	(internet* or world wide web or worldwide web or web site? or website? or web portal? or blog* or web log* or bulletin board? or bulletinboard? or message board? or messageboard? or forum? or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page? or webpage? or chat room? og chatroom?).ab,ti.	1172
73	or/67-72	14137
74	66 and 73	341
75	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or consumer or medical or scien*) adj2 information)).ab,ti.	70
76	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 risk information).ab,ti.	1
77	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or medical or scien*) adj2 (claim? or statement? or message?))).ab,ti.	9
78	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 ((health or medical or scien*) adj2 (paper? or article? or report? or literature? or journal? or periodical? or research))).ab,ti.	267
79	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) adj2 information) or pamphlet? or leaflet? or booklet? or brochure?)).ab,ti.	662
80	(((critical* adj read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) adj7 (internet* or world wide web or worldwide web or web site? or website? or web portal? or blog* or web log* or bulletin board? or bulletinboard? or message board? or messageboard? or forum? or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page? or webpage? or chat room? or chatroom?)).ab,ti.	123

81	or/36,74-80	5341
82	1 and 20 and 81	17
83	10 and 81	207
84	82 or 83	220
85	randomized controlled trials/	1649
86	comparative study/	4586
87	clinical trials/	1721
88	program evaluation/	1805
89	follow up studies/	1103
90	longitudinal studies/	273
91	prospective studies/	693
92	pilot projects/	1054
93	(randomis* or randomiz* or randomly or random allocat*).ti,ab.	12077
94	(group? and (random* or between* or control* or intervent*)).ab.	16823
95	(multicenter or multi center or multicentre or multi centre).ti. or trial.ti,ab.	8138
96	(intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) adj5 (study or studies or design)) or (before adj5 after) or (pre adj5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect? or effectiveness or impact or time series or time point? or repeated measur*).ti,ab.	99330
97	or/85-96	104844
98	84 and 97	98

Database: Cochrane Central Register of Controlled Trials (CENTRAL) via The Cochrane Library (Wiley)
Date: 25.02.2014
Hits: 307

ID	Search	Hits
#1	(age* near/2 (11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)):ti,ab	17507
#2	MeSH descriptor: [Schools] this term only	866
#3	MeSH descriptor: [School Health Services] this term only	797
#4	MeSH descriptor: [Students] this term only	1146
#5	(schoolchild* or school-child* or school student* or pupil?):ti,ab	4541
#6	((middle or secondary or high) next school?):ti,ab	511
#7	((six or sixth or seven* or eight? or nine or ninth or ten or tenth or eleven* or twelve or twelfth) near/3 grade*):ti,ab	763
#8	((grade*) next (6 or 7 or 8 or 9 or 10 or 11 or 12)):ti,ab	259
#9	((year*) next (6 or 7 or 8 or 9 or 10 or 11 or 12 or 13)):ti,ab	1363
#10	#2 or #3 or #4 or #5 or #6 or #7 or #8 or #9	7968
#11	MeSH descriptor: [Education] this term only	447
#12	MeSH descriptor: [Health Education] this term only	2750
#13	MeSH descriptor: [Health Education, Dental] this term only	210
#14	MeSH descriptor: [Health Fairs] this term only	6
#15	MeSH descriptor: [Sex Education] this term only	186
#16	MeSH descriptor: [Patient Education as Topic] this term only	6065
#17	MeSH descriptor: [Education, Distance] this term only	65
#18	MeSH descriptor: [Education, Nonprofessional] this term only	5
#19	MeSH descriptor: [Education, Special] this term only	128
#20	MeSH descriptor: [Competency-Based Education] this term only	60
#21	MeSH descriptor: [Programmed Instruction as Topic] explode all trees	921
#22	MeSH descriptor: [Teaching] this term only	1208
#23	MeSH descriptor: [Curriculum] this term only	845
#24	MeSH descriptor: [Teaching Materials] explode all trees	2710
#25	MeSH descriptor: [Educational Technology] explode all trees	2305
#26	Any MeSH descriptor with qualifier(s): [Education - ED]	4709
#27	(educat* or train* or teach* or workshop? or work-shop? or seminar? or course? or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson? or taught or module?):ti,ab	75764
#28	#11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27	80773

#29	MeSH descriptor: [Consumer Health Information] explode all trees	125
#30	MeSH descriptor: [Information Services] this term only	97
#31	MeSH descriptor: [Research] this term only	276
#32	MeSH descriptor: [Empirical Research] this term only	42
#33	MeSH descriptor: [Research Design] this term only	5071
#34	MeSH descriptor: [Research Report] this term only	14
#35	MeSH descriptor: [Statistics as Topic] this term only	2370
#36	MeSH descriptor: [Periodicals as Topic] this term only	77
#37	MeSH descriptor: [Epidemiology] this term only	15
#38	MeSH descriptor: [Pharmacoepidemiology] this term only	13
#39	MeSH descriptor: [Pamphlets] this term only	572
#40	MeSH descriptor: [Newspapers] this term only	17
#41	MeSH descriptor: [Mass Media] this term only	161
#42	MeSH descriptor: [Television] this term only	221
#43	MeSH descriptor: [Radio] this term only	21
#44	MeSH descriptor: [Internet] explode all trees	1525
#45	#29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44	10240
#46	MeSH descriptor: [Information Seeking Behavior] this term only	10
#47	MeSH descriptor: [Problem Solving] this term only	1102
#48	MeSH descriptor: [Problem-Based Learning] this term only	180
#49	MeSH descriptor: [Data Interpretation, Statistical] this term only	1548
#50	MeSH descriptor: [Evidence-Based Practice] explode all trees	2048
#51	#46 or #47 or #48 or #49 or #50	4847
#52	#45 and #51	508
#53	MeSH descriptor: [Science] this term only	20
#54	MeSH descriptor: [Information Literacy] explode all trees	75
#55	MeSH descriptor: [Judgment] this term only	452
#56	MeSH descriptor: [Decision Making] this term only	1470
#57	MeSH descriptor: [Thinking] this term only	351
#58	#52 or #53 or #54 or #55 or #56 or #57	2792
#59	((health or information or mathematical or quantitative or science or scientific* or media) near/2 (literacy or literate? or illiteracy or illiterate?)):ti,ab	192
#60	(information near/2 competen*):ti,ab	9
#61	numeracy:ti,ab	37
#62	(scientific near/2 (skill? or think* or reason*)):ti,ab	19
#63	(critical near/2 (think* or reason*)):ti,ab	58

#64	evidence-based:ti,ab	3896
#65	#59 or #60 or #61 or #62 or #63 or #64	4187
#66	health next litera*	546
#67	medical next literacy	2
#68	(health and literacy):ti	71
#69	(functional health literacy):ti,ab,kw	20
#70	low-litera*:ti	30
#71	litera*:ti	2596
#72	illitera*:ti	12
#73	MeSH descriptor: [Reading] this term only	629
#74	MeSH descriptor: [Comprehension] this term only	303
#75	#66 or 70 or #68 or #69 or #70 or #71 or #72 or #73 or #74	51444
#76	MeSH descriptor: [Health Promotion] this term only	3328
#77	MeSH descriptor: [Health Education] this term only	2750
#78	MeSH descriptor: [Patient Education as Topic] this term only	6065
#79	MeSH descriptor: [Communication Barriers] this term only	76
#80	MeSH descriptor: [Communication] this term only	1262
#81	MeSH descriptor: [Health Knowledge, Attitudes, Practice] this term only	3307
#82	MeSH descriptor: [Attitude to Health] this term only	2448
#83	MeSH descriptor: [Comprehension] this term only	303
#84	MeSH descriptor: [Educational Status] this term only	1032
#85	#83 and #84	28
#86	(family and literacy):ti	4
#87	drug labeling:ti,ab,kw	13289
#88	MeSH descriptor: [Prescription Drugs] this term only	68
#89	comprehension:ti,ab,kw	954
#90	(#87 or #88) and #89	19
#91	((cancer or diabetes or genetics) and (literacy or comprehension)):ti	31
#92	(limited and (educational status or (educational and status) or literacy)):ti,ab,kw	178
#93	(patient* and (educational status or (educational and status) or literacy)):ti,ab,kw	1179
#94	(patient* and (comprehension or understanding)):ti,ab,kw	4580
#95	#76 or #77 or #78 or #79 or #80 or #81 or #82 or #85 or #86 or #90 or #91 or #92 or #93 or #94	20019
#96	#75 and #95	1909
#97	#65 or #96	5908
#98	((health or consumer or medical or scien*) near/2 information):ti,ab	1106
#99	risk next information:ti,ab	156
#100	((health or medical or scien*) near/2 (claim? or statement? or message?)):ti,ab	158

#101	((health or medical or scien*) near/2 (paper? or article? or report? or literature? or journal? or periodical? or research)):ti,ab	5842
#102	(media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) near/2 information) or pamphlet? or leaflet? or booklet? or brochure?):ti,ab	60206
#103	(internet* or (world next wide next web) or (worldwide next web) or (web next site?) or website? or (web next portal?) or blog* or (web next log*) or (bulletin next board?) or bulletinboard? or (message next board?) or messageboard? or forum? or ehealth or e-health or (electronic next health) or weblog* or moblog* or vlog* or (video next blog*) or microblog* or wiki* or (web next page?) or webpage? or (chat next room?) or chatroom?):ti,ab	2758
#104	#98 or #99 or #100 or #101 or #102 or #103	68647
#105	#97 and #104	1065
#106	(((critical* next read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) near/7 ((health or consumer or medical or scien*) near/2 information)):ti,ab	123
#107	(((critical* next read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) near/7 risk information):ti,ab	1024
#108	(((critical* next read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) near/7 ((health or medical or scien*) near/2 (claim? or statement? or message?))):ti,ab	12
#109	(((critical* next read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) near/7 ((health or medical or scien*) near/2 (paper? or article? or report? or literature? or journal? or periodical? or research))):ti,ab	308
#110	(((critical* next read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) near/7 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine? or ((print* or written) near/2 information) or pamphlet? or leaflet? or booklet? or brochure?)):ti,ab	2426
#111	(((critical* next read*) or apprais* or evaluat* or understand or assess* or judg* or interpret*) near/7 (internet* or (world next wide next web) or (worldwide next web) or (web next site?) or website? or (web next portal?) or blog* or (web next log*) or (bulletin next board?) or bulletinboard? or (message next board?) or messageboard? or forum? or ehealth or e-health or (electronic next health) or weblog* or moblog* or vlog* or (video next blog*) or microblog* or wiki* or (web next page?) or webpage? or (chat next room?) or chatroom?)):ti,ab	470
#112	#58 or #105 or #106 or #107 or #108 or #109 or #110 or #111	7817
#113	#1 and #28 and #112	111
#114	#10 and #112	261
#115	#113 or #114 in Trials	307

Database: Education Resource Information Center (ERIC) via EBSCOhost
Date: 25.02.2014
Hits: 6,261

#	Query	Results
S1	DE "Middle Schools"	23,711
S2	DE "Secondary Schools"	5,123
S3	DE "High Schools"	50,385
S4	DE "Vocational High Schools"	314
S5	DE "Junior High Schools"	16,709
S6	DE "Elementary Secondary Education"	213,766
S7	DE "Elementary Education"	96,730
S8	DE "Secondary Education"	105,381
S9	DE "College Preparation"	4,551
S10	DE "Elementary School Curriculum"	2,666
S11	DE "Elementary School Science"	9,633
S12	DE "Secondary School Curriculum"	3,291
S13	DE "Secondary School Science"	17,447
S14	DE "Intermediate Grades"	14,944
S15	DE "Grade 10" OR DE "Grade 11" OR DE "Grade 12" OR DE "Grade 6" OR DE "Grade 7" OR DE "Grade 8" OR DE "Grade 9"	28,761
S16	Limiters - Educational Level: Elementary Education, Elementary Secondary Education, Grade 6, Grade 7, Grade 8, Grade 9, Grade 10, Grade 11, Grade 12, High Schools, Intermediate Grades, Junior High Schools, Middle Schools, Secondary Education	478,469
S17	TI (schoolchild* or school-child* or school student# or pupil#) OR AB (schoolchild* or school-child* or school student# or pupil#)	125,975
S18	TI ((middle or secondary or high) N0 School#) OR AB ((middle or secondary or high) N0 School#)	122,068
S19	TI ((six or sixth or seven* or eight# or nine or ninth or ten or tenth or eleven* or twelve or twelfth) N0 grade#) OR AB ((six or sixth or seven* or eight# or nine or ninth or ten or tenth or eleven* or twelve or twelfth) N0 grade#)	28,840
S20	TI ((grade# 6 or year 6 or grade# 7 or year 7 or grade# 8 or year 8 or grade# 9 or year 9 or grade# 10 or year 10 or grade# 11 or year 11 or grade# 12 or year 12 or year 13)) OR AB ((grade# 6 or year 6 or grade# 7 or year 7 or grade# 8 or year 8 or grade# 9 or year 9 or grade# 10	82,245

	or year 10 or grade# 11 or year 11 or grade# 12 or year 12 or year 13))	
S21	TI (age# N2 ("11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)) OR AB (age# N2 ("11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen))	19,229
S22	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21	576,957
S23	DE "Health Education"	10,275
S24	DE "Comprehensive School Health Education"	850
S25	DE "Alcohol Education"	1,324
S26	DE "Drug Education"	2,529
S27	DE "Physical Education"	8,857
S28	DE "Sex Education"	3,069
S29	DE "Patient Education"	545
S30	DE "Nutrition Instruction"	2,075
S31	DE "Foods Instruction"	804
S32	DE "Home Economics"	2,058
S33	DE "School Health Services"	1,940
S34	DE "Controversial Issues (Course Content)"	2,185
S35	DE "Health Materials"	472
S36	DE "Evidence"	4,464
S37	DE "Information Services"	5,119
S38	DE "Community Information Services"	585
S39	DE "Information Dissemination"	10,236
S40	DE "Information Sources"	8,088
S41	DE "Information Utilization"	3,588
S42	DE "Evaluation Utilization"	1,623
S43	DE "Research Utilization"	5,098
S44	DE "Epidemiology"	775
S45	DE "Research"	15,219

S46	DE "Medical Research"	1,398
S47	DE "Scientific Research"	5,738
S48	DE "Research Design"	7,750
S49	DE "Research Reports"	5,562
S50	DE "Scientific and Technical Information"	1,355
S51	DE "Science Materials"	2,392
S52	DE "Sciences"	5,422
S53	DE "Statistics"	4,125
S54	DE "Periodicals"	5,746
S55	DE "Electronic Journals"	693
S56	DE "Journal Articles"	1,514
S57	DE "Pamphlets"	449
S58	DE "Mass Media"	7,070
S59	DE "Newspapers"	4,176
S60	DE "News Media"	2,963
S61	DE "Newsletters"	1,626
S62	DE "Radio"	2,159
S63	DE "Educational Radio"	1,169
S64	DE "Television"	5,036
S65	DE "Educational Television"	6,290
S66	DE "Public Television"	929
S67	DE "Internet"	17,854
S68	DE "Web 2.0 Technologies"	491
S69	DE "Printed Materials"	1,118
S70	DE "Reference Materials"	4,313
S71	S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67 OR S68 OR S69 OR S70	155,911
S72	DE "Credibility"	1,501

S73	DE "Critical Thinking"	10,903
S74	DE "Evaluative Thinking"	1,539
S75	DE "Thinking Skills"	10,942
S76	DE "Logical Thinking"	3,689
S77	DE "Critical Literacy"	113
S78	DE "Critical Viewing"	374
S79	DE "Critical Reading"	1,988
S80	DE "Criticism"	2,977
S81	DE "Information Literacy"	2,236
S82	DE "Information Skills"	790
S83	DE "Media Literacy"	887
S84	DE "Numeracy"	1,813
S85	DE "Data Interpretation"	1,440
S86	DE "Scientific Literacy"	2,774
S87	DE "Science Process Skills"	1,734
S88	DE "Scientific Principles"	3,348
S89	DE "Problem Solving"	31,360
S90	DE "Problem Based Learning"	2,367
S91	DE "Decision Making Skills"	2,552
S92	S72 OR S73 OR S74 OR S75 OR S76 OR S77 OR S78 OR S79 OR S80 OR S81 OR S82 OR S83 OR S84 OR S85 OR S86 OR S87 OR S88 OR S89 OR S90 OR S91	72,597
S93	TI ((health or information or mathematical or quantitative or science or scientific* or media) N2 (literacy or literate* or illiteracy or illiterate*)) OR AB ((health or information or mathematical or quantitative or science or scientific* or media) N2 (literacy or literate* or illiteracy or illiterate*))	5,611
S94	TI (information N2 competen*) OR AB (information N2 competen*)	495
S95	TI numeracy OR AB numeracy	1,613
S96	TI ((scientific N2 (skill* or think* or reason*))) OR AB ((scientific N2 (skill* or think* or reason*)))	1,176
S97	TI evidence-based OR AB evidence-based	3,635

S98	TI (critical N2 (think* or reason*)) OR AB (critical N2 (think* or reason*))	8,617
S99	S93 OR S94 OR S95 OR S96 OR S97 OR S98	20,392
S100	TI ((health or consumer or medical or scien*) N2 information) OR AB ((health or consumer or medical or scien*) N2 information)	7,281
S101	TI risk information OR AB risk information	548
S102	TI (((health or medical or scien*) N2 (claim# or statement* or message*))) OR AB (((health or medical or scien*) N2 (claim# or statement* or message*)))	641
S103	TI (((health or medical or scien*) N0 (paper# or article* or report# or literature* or journal# or periodical# or research))) OR AB (((health or medical or scien*) N0 (paper# or article* or report# or literature* or journal# or periodical# or research)))	6,072
S104	TI ((media* or television* or tv or radio or broadcast* or broadsides or news* or magazine# or magasine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#)) OR AB ((media* or television* or tv or radio* or broadcast* or broadsides or news* or magazine# or magasine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#))	107,949
S105	TI ((internet* or world wide web or worldwide web or web site# or website# or web page# or webpage# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum* or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or chat room# or chatroom#)) OR AB ((internet* or world wide web or worldwide web or web site# or website# or web page# or webpage# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum* or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or chat room# or chatroom#))	43,315
S106	S100 OR S101 OR S102 OR S103 OR S104 OR S105	156,363
S107	S99 AND S106	4,342
S108	TI (((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or consumer or medical or scien*) N2 information)) OR AB (((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or consumer or medical or scien*) N2 information))	526
S109	TI (((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 risk information) OR AB (((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 risk information)	68
S110	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or medical or scien*) N0 (claim# or statement* or message*)))) OR AB ((((critical*	31

	N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or medical or scien*) N0 (claim# or statement* or message*))))	
S111	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or medical or scien*) N0 (paper# or article* or report# or literature* or journal# or periodical# or research)))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or medical or scien*) N0 (paper# or article* or report# or literature* or journal# or periodical# or research))))	491
S112	TI (((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (media* or television* or tv or radio or broadcast* or broadsides or news* or magazine# or magasine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#)) OR AB (((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (media* or television* or tv or radio or broadcast* or broadsides or news* or magazine# or magasine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#))	9,622
S113	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (internet* or world wide web or worldwide web or web site# or website# or web page# or web page# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum* or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or chat room# or chatroom#))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (internet* or world wide web or worldwide web or web site# or website# or web page# or web page# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum* or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or chat room# or chatroom#)))	2,941
S114	(S71 AND S92) OR S107 OR S108 OR S109 OR S110 OR S111 OR S112 OR S113	24,223
S115	S22 AND S114	10,439
S116	DE "Experimental Groups"	2,829
S117	DE "Quasiexperimental Design"	1,088
S118	DE "Control Groups"	6,174
S119	DE "Matched Groups"	560
S120	DE "Evaluation Research"	2,055
S121	DE "Pretests Posttests"	7,707
S122	DE "Pilot Projects"	5,264
S123	DE "Comparative Analysis"	59,571

S124	TI ((randomis* or randomiz* or randomly or random allocat*)) OR AB ((randomis* or randomiz* or randomly or random allocat*))	14,641
S125	AB (group# and (random* or between* or control* or intervent*))	71,416
S126	TI ((multicenter or multi center or multicentre or multi centre)) OR TI trial OR AB trial	10,119
S127	TI ((intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) N5 (study or studies or design)) or (before N5 after) or (pre N5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect* or effectiveness or impact or time series or time point* or repeated measur*)) OR AB ((intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) N5 (study or studies or design)) or (before N5 after) or (pre N5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect* or effectiveness or impact or time series or time point* or repeated measur*))	551,243
S128	S116 OR S117 OR S118 OR S119 OR S120 OR S121 OR S122 OR S123 OR S124 OR S125 OR S126 OR S127	596,536
S129	S115 AND S128	6,261

Database: Teacher Reference Center (TRC) via EBSCOhost
Date: 25.02.2014
Hits: 290

#	Query	Results
S1	ZU "secondary schools"	72
S2	ZU "middle schools"	1,327
S3	ZU "high schools"	3,378
S4	ZU "junior high schools"	98
S5	ZU "special education schools"	53
S6	ZU "elementary schools"	2,203
S7	ZU "secondary education"	7,881
S8	ZU "elementary education"	4,007
S9	ZU "middle school students"	1,154
S10	ZU "high school students"	2,595
S11	ZU "secondary school students"	5
S12	ZU "sixth grade (education)"	106
S13	ZU "seventh grade (education)"	68
S14	ZU "eighth grade (education)"	105
S15	ZU "ninth grade (education)"	95
S16	ZU "tenth grade (education)"	31
S17	ZU "eleventh grade (education)"	17
S18	ZU "twelfth grade (education)"	25
S19	TI (schoolchild* or school-child* or school student* or pupil#) OR AB (schoolchild* or school-child* or school student* or pupil#)	48,649
S20	TI ((middle or secondary or high) N0 school#) OR AB ((middle or secondary or high) N0 school#)	38,416
S21	TI ((six or sixth or seven* or eight# or nine or ninth or ten or tenth or eleven* or twelve or twelfth) N0 grade*) OR AB ((six or sixth or seven* or eight# or nine or ninth or ten or tenth or eleven* or twelve or twelfth) N0 grade*)	4,375
S22	TI ((grade# 6 or year 6 or grade# 7 or year 7 or grade# 8 or year 8 or grade# 9 or year 9 or grade# 10 or year 10 or grade# 11 or year 11 or grade# 12 or year 12 or year 13)) OR AB ((grade# 6 or year 6 or grade# 7 or year 7 or grade# 8 or year 8 or grade# 9 or year 9 or grade# 10	16,857

	or year 10 or grade# 11 or year 11 or grade# 12 or year 12 or year 13))	
S23	TI (age# N2 ("11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)) OR AB (age# N2 ("11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen))	2,569
S24	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23	96,294
S25	ZU "health education"	767
S26	ZU "health education (middle school)"	18
S27	ZU "health education (secondary)"	41
S28	ZU "dental health education"	16
S29	ZU "sex education"	558
S30	ZU "health fairs"	1
S31	ZU "physical education"	1,874
S32	ZU "patient education"	47
S33	ZU "home economics"	117
S34	ZU "school health services"	412
S35	ZU "evidence"	78
S36	ZU "information services"	1,073
S37	ZU "community information services"	7
S38	ZU "information dissemination"	25
S39	ZU "information resources"	1,692
S40	ZU "information retrieval"	328
S41	ZU "information science"	683
S42	ZU "information-seeking behavior"	31
S43	ZU "information-seeking strategies"	16
S44	ZU "evaluation utilization"	6
S45	ZU "research -- evaluation"	538
S46	ZU "research -- methodology"	1,093
S47	ZU "research use"	5

S48	ZU "scientific experimentation"	1,592
S49	ZU "scientific knowledge"	138
S50	ZU "epidemiology"	313
S51	ZU "medical research"	553
S52	ZU "medical sciences"	212
S53	ZU "science"	3,194
S54	ZU "science classrooms & equipment"	100
S55	ZU "science experiments"	256
S56	ZU "science in literature"	40
S57	ZU "science in mass media"	14
S58	ZU "science journalism"	15
S59	ZU "science on television"	4
S60	ZU "science projects"	455
S61	ZU "science students"	583
S62	ZU "science television programs"	13
S63	ZU "scientific apparatus & instruments"	238
S64	ZU "scientific development"	38
S65	ZU "scientific experimentation"	1,592
S66	(ZU "statistics")	2,215
S67	ZU "periodicals"	3,601
S68	ZU "pamphlets"	206
S69	ZU "electronic journals"	349
S70	ZU "mass media"	795
S71	ZU "newspapers"	250
S72	ZU "newspapers in education"	61
S73	ZU "newsletters"	163
S74	ZU "radio"	46
S75	ZU "educational television programs"	126
S76	ZU "public television"	22

S77	ZU "internet"	3,210
S78	ZU "web 2.0"	262
S79	ZU "reference sources"	460
S80	ZU "web portals"	151
S81	S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67 OR S68 OR S69 OR S70 OR S71 OR S72 OR S73 OR S74 OR S75 OR S76 OR S77 OR S78 OR S79 OR S80	27,022
S82	ZU "health literacy"	161
S83	ZU "health literacy -- research"	4
S84	ZU "information literacy"	315
S85	ZU "information literacy -- research"	4
S86	ZU "information literacy -- study & teaching"	35
S87	ZU "critical literacy"	25
S88	ZU "judgment"	148
S89	ZU "decision making"	1,740
S90	ZU "problem solving"	2,130
S91	ZU "problem-based learning"	675
S92	ZU "critical thinking"	1,246
S93	ZU "critical thinking -- study & teaching"	218
S94	ZU "critical thinking -- research"	14
S95	ZU "criticism"	1,036
S96	ZU "information skills"	14
S97	ZU "media literacy"	185
S98	ZU "numeracy"	462
S99	ZU "numeracy -- study & teaching"	65
S100	ZU "scientific literacy"	35
S101	ZU "problem solving -- research"	30
S102	ZU "problem solving -- study & teaching"	158

S103	ZU "decision making in adolescence"	11
S104	ZU "decision making in children"	16
S105	ZU "decision making in children"	16
S106	ZU "decision making"	1,740
S107	(ZU "critical thinking in adolescence")	9
S108	(ZU "critical thinking in children")	49
S109	(ZU "critical thinking in children -- study & teaching")	9
S110	(ZU "critical analysis")	49
S111	S82 OR S83 OR S84 OR S85 OR S86 OR S87 OR S88 OR S89 OR S90 OR S91 OR S92 OR S93 OR S94 OR S95 OR S96 OR S97 OR S98 OR S99 OR S100 OR S101 OR S102 OR S103 OR S104 OR S105 OR S106 OR S107 OR S108 OR S109 OR S110	8,261
S112	S81 AND S111	522
S113	TI (((health or information or mathematical or quantitative or science or scientific* or media) N2 (literacy or literate# or illiteracy or illiterate#))) OR AB (((health or information or mathematical or quantitative or science or scientific* or media) N2 (literacy or literate# or illiteracy or illiterate#)))	1,841
S114	TI (information N2 competen*) OR AB (information N2 competen*)	57
S115	TI numeracy OR AB numeracy	1,327
S116	TI ((scientific N2 (skill# or think* or reason*))) OR AB ((scientific N2 (skill# or think* or reason*)))	330
S117	TI ((critical N2 (think* or reason*))) OR AB ((critical N2 (think* or reason*)))	2,074
S118	TI evidence-based OR AB evidence-based	927
S119	S113 OR S114 OR S115 OR S116 OR S117 OR S118	6,379
S120	TI (((health or consumer or medical or scien*) N2 information)) OR AB (((health or consumer or medical or scien*) N2 information))	2,842
S121	TI risk information OR AB risk information	145
S122	TI (((health or medical or scien*) N2 (claim# or statement# or message#))) OR AB (((health or medical or scien*) N2 (claim# or statement# or message#)))	270
S123	TI (((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research))) OR AB (((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research)))	7,416
S124	TI ((media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or	50,578

	((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#)) OR AB ((media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#))	
S125	TI ((internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#)) OR AB ((internet* or world wide web or worldwide web or web site# or website# or web portal* or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#))	32,352
S126	S120 OR S121 OR S122 OR S123 OR S124 OR S125	87,395
S127	S119 AND S126	1,185
S128	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or consumer or medical or scien*) N2 information))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or consumer or medical or scien*) N2 information)))	126
S129	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 risk information)) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 risk information))	10
S130	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or medical or scien*) N2 (claim# or statement# or message#)))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or medical or scien*) N2 (claim# or statement# or message#))))	19
S131	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research)))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 ((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research))))	492
S132	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#)))	1,179

S133	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#)))	1,223
S134	S112 OR S127 OR S128 OR S129 OR S130 OR S131 OR S132 OR S133	4,443
S135	S24 AND S134	628
S136	ZU "randomized controlled trials"	171
S137	ZU "clinical trials"	270
S138	ZU "pre-tests & post-tests"	305
S139	ZU "crossover trials"	35
S140	ZU "time series analysis"	20
S141	ZU "pilot projects"	284
S142	ZU "repeated measures design"	163
S143	TI ((randomis* or randomiz* or randomly or random allocat*)) OR AB ((randomis* or randomiz* or randomly or random allocat*))	1,634
S144	AB (group? and (random* or between* or control* or intervent*))	4,199
S145	TI ((multicenter or multi center or multicentre or multi centre)) OR TI trial OR AB trial	2,749
S146	TI ((intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) N5 (study or studies or design)) or (before N5 after) or (pre N5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect* or effectiveness or impact or time series or time point* or repeated measur*)) OR AB ((intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) N5 (study or studies or design)) or (before N5 after) or (pre N5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect* or effectiveness or impact or time series or time point* or repeated measur*))	114,166
S147	S136 OR S137 OR S138 OR S139 OR S140 OR S141 OR S142 OR S143 OR S144 OR S145 OR S146	118,055

S148	S135 AND S147	290

Database: Library, Information Science & Technology Abstracts (LISTA) via EBSCOhost
Date: 25.02.2014
Hits: 481/28 (509)

#	Query	Results
S1	TI ((age# N2 (11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen))) OR AB ((age# N2 (11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)))	948
S2	DE "SCHOOLS"	1,569
S3	DE "STUDENTS"	5,351
S4	TI ((schoolchild* or school-child* or school student* or pupil#)) OR AB ((schoolchild* or school-child* or school student* or pupil#))	3,045
S5	TI (((middle or secondary or high) N0 school#)) OR AB (((middle or secondary or high) N0 school#))	6,093
S6	TI (((six or sixth or seven* or eight# or nine or ninth or ten or tenth or eleven* or twelve or twelfth) N3 grade*)) OR AB (((six or sixth or seven* or eight# or nine or ninth or ten or tenth or eleven* or twelve or twelfth) N3 grade*))	5,660
S7	TI ((grade# 6 or grade# 7 or grade# 8 or grade# 9 or grade# 10 or grade# 11 or grade# 12)) OR AB ((grade# 6 or grade# 7 or grade# 8 or grade# 9 or grade# 10 or grade# 11 or grade# 12)) OR TI ((year# 6 or year# 7 or year# 8 or year# 9 or year# 10 or year# 11 or year# 12 or year# 13)) OR AB ((year# 6 or year# 7 or year# 8 or year# 9 or year# 10 or year# 11 or year# 12 or year# 13))	549
S8	S2 OR S3 OR S4 OR S5 OR S6 OR S7	19,758
S9	DE "EDUCATION"	13,304
S10	DE "CURRICULA (Courses of study)"	2,246
S11	DE "DISTANCE education"	2,216
S12	DE "LITERACY education"	43
S13	DE "USER education"	134
S14	DE "COMPUTER assisted instruction"	1,649

S15	DE "EDUCATIONAL technology"	4,289
S16	DE "AUDIOVISUAL education"	145
S17	TI ((educat* or train* or teach* or workshop# or work-shop# or seminar# or course# or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson# or taught or module#)) OR AB ((educat* or train* or teach* or workshop# or work-shop# or seminar# or course# or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson# or taught or module#))	169,827
S18	S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17	173,208
S19	DE "INFORMATION services"	17,166
S20	DE "RESEARCH"	16,897
S21	DE "COMPUTER assisted research"	46
S22	DE "EXPERIMENTAL design"	215
S23	DE "EXPERIMENTS"	223
S24	DE "INFORMATION science -- Research"	432
S25	DE "INFORMATION resources -- Research"	82
S26	DE "INFORMATION resources -- Use studies"	121
S27	DE "INTERNET research"	498
S28	DE "OPERATIONS research"	128
S29	DE "QUALITATIVE research"	653
S30	DE "QUANTITATIVE research"	432
S31	DE "NEWSPAPERS"	1,631
S32	DE "SERIAL publications"	3,174
S33	DE "ELECTRONIC newspapers"	238
S34	DE "PERIODICALS"	10,797
S35	DE "PAMPHLETS"	243

S36	DE "BROCHURES"	17
S37	DE "DIGITAL media"	1,756
S38	DE "INTERNET"	17,430
S39	DE "INTERNET in education"	3,351
S40	DE "VIRTUAL communities"	631
S41	DE "MASS media"	2,362
S42	DE "WEB 2.0"	1,752
S43	DE "BLOGS"	2,766
S44	DE "WIKIS (Computer science)"	716
S45	DE "AUDIOVISUAL materials"	2,268
S46	S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45	76,678
S47	DE "PROBLEM solving"	899
S48	DE "SEARCHING behavior"	181
S49	S47 OR S48	1,074
S50	S46 AND S49	155
S51	DE "ELECTRONIC information resource literacy"	230
S52	DE "HEALTH literacy"	94
S53	DE "MEDIA literacy"	259
S54	DE "INTERNET literacy"	127
S55	DE "LITERACY"	2,337
S56	DE "COMPUTER literacy"	489
S57	DE "INFORMATION literacy"	4,295
S58	DE "VISUAL literacy"	69

S59	S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58	7,346
S60	TI (((health or information or mathematical or quantitative or science or scientific* or media) N2 (literacy or literate# or illiteracy or illiterate#))) OR AB (((health or information or mathematical or quantitative or science or scientific* or media) N2 (literacy or literate# or illiteracy or illiterate#)))	5,841
S61	TI (information N2 competen*) OR AB (information N2 competen*)	597
S62	TI numeracy OR AB numeracy	84
S63	TI ((scientific N2 (skill# or think* or reason*))) OR AB ((scientific N2 (skill# or think* or reason*)))	92
S64	TI ((critical N2 (think* or reason*))) OR AB ((critical N2 (think* or reason*)))	713
S65	TI evidence-based OR AB evidence-based	1,111
S66	S60 OR S61 OR S62 OR S63 OR S64 OR S65	7,894
S67	TI (((health or consumer or medical or scien*) N2 information)) OR AB (((health or consumer or medical or scien*) N2 information))	23,504
S68	TI risk information OR AB risk information	35
S69	TI (((health or medical or scien*) N2 (claim# or statement# or message#))) OR AB (((health or medical or scien*) N2 (claim# or statement# or message#)))	208
S70	TI (((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research))) OR AB (((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research)))	13,183
S71	TI ((media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#)) OR AB ((media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#))	90,974
S72	TI ((internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#)) OR AB ((internet* or world wide web or worldwide web or web site# or website# or web portal* or blog* or web log* or bulletin board# or	77,497

	bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#))	
S73	S67 OR S68 OR S69 OR S70 OR S71 OR S72	183,040
S74	S66 AND S73	2,608
S75	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or consumer or medical or scien*) N2 information))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or consumer or medical or scien*) N2 information)))	879
S76	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 risk information)) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 risk information))	5
S77	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or medical or scien*) N2 (claim# or statement# or message#)))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or medical or scien*) N2 (claim# or statement# or message#))))	6
S78	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research)))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research))))	576
S79	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#)))	2,087
S80	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 (internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message	2,876

	board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 (internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#)))	
S81	S59 OR S74 OR S75 OR S76 OR S77 OR S78 OR S79 OR S80	14,468
S82	S1 AND S18 AND S81	45
S83	S8 AND S81	1,086
S84	DE "EXPERIMENTAL design"	215
S85	TI ((randomis* or randomiz* or randomly or random allocat*)) OR AB ((randomis* or randomiz* or randomly or random allocat*))	1,312
S86	AB (group? and (random* or between* or control* or intervent*))	2,825
S87	TI ((multicenter or multi center or multicentre or multi centre)) OR TI trial OR AB trial	4,375
S88	TI ((intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) N5 (study or studies or design)) or (before N5 after) or (pre N5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect* or effectiveness or impact or time series or time point* or repeated measur*)) OR AB ((intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) N5 (study or studies or design)) or (before N5 after) or (pre N5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect* or effectiveness or impact or time series or time point* or repeated measur*))	102,266
S89	S84 OR S85 OR S86 OR S87 OR S88	107,367
S90	S82 AND S89	28
S91	S83 AND S89	481

Database: Cumulative Index to Nursing and Allied Health Literature (Cinahl) via EBSCOhost
Date: 25.02.2014
Hits: 884

#	Query	Results
S1	TI ((age# N2 (11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen))) OR AB ((age# N2 (11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)))	38,922
S2	(MH "Schools")	4,262
S3	(MH "Schools, Middle")	913
S4	(MH "Schools, Secondary")	2,492
S5	(MH "Schools, Special")	569
S6	(MH "School Health Services")	4,989
S7	(MH "Students")	3,571
S8	(MH "Students, High School")	4,711
S9	(MH "Students, Middle School")	1,778
S10	TI ((schoolchild* or school-child* or school student* or pupil#)) OR AB ((schoolchild* or school-child* or school student* or pupil#))	8,137
S11	TI (((middle or secondary or high) N0 school#)) OR AB (((middle or secondary or high) N0 school#))	9,524
S12	TI (((six or sixth or seven* or eight# or nine or ninth or ten or tenth or eleven* or twelve or twelfth) N3 grade*)) OR AB (((six or sixth or seven* or eight# or nine or ninth or ten or tenth or eleven* or twelve or twelfth) N3 grade*))	1,993
S13	TI ((grade# 6 or grade# 7 or grade# 8 or grade# 9 or grade# 10 or grade# 11 or grade# 12)) OR AB ((grade# 6 or grade# 7 or grade# 8 or grade# 9 or grade# 10 or grade# 11 or grade# 12)) OR TI ((year# 6 or year# 7 or year# 8 or year# 9 or year# 10 or year# 11 or year# 12 or year# 13)) OR AB ((year# 6 or year# 7 or year# 8 or year# 9 or year# 10 or year# 11 or year# 12 or year# 13))	2,876
S14	S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13	32,255

S15	(MH "Education")	4,081
S16	(MH "Health Education")	13,269
S17	(MH "School Health Education")	3,565
S18	(MH "Dental Health Education")	225
S19	(MH "Health Fairs")	388
S20	(MH "Sex Education")	2,743
S21	(MH "Nutrition Education")	4,802
S22	(MH "Patient Education")	37,530
S23	(MH "Education, Non-Traditional")	5,674
S24	(MH "Education, Special")	2,626
S25	(MH "Education, Nonprofessional")	25
S26	(MH "Education, Competency-Based")	1,538
S27	(MH "Curriculum+")	19,589
S28	(MH "Programmed Instruction+")	4,671
S29	(MH "Teaching")	3,122
S30	(MH "Teaching Materials+")	64,967
S31	(MH "Educational Technology")	762
S32	TI ((educat* or train* or teach* or workshop# or work-shop# or seminar# or course# or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson# or taught or module#)) OR AB ((educat* or train* or teach* or workshop# or work-shop# or seminar# or course# or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson# or taught or module#))	373,268
S33	S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32	459,255
S34	(MH "Health Information+")	13,678
S35	(MH "Information Services")	5,104

S36	(MH "Research")	18,195
S37	(MH "Empirical Research")	1,547
S38	(MH "Study Design")	9,438
S39	(MH "Statistics")	3,790
S40	(MH "Epidemiology")	2,425
S41	(MH "Newspapers")	1,795
S42	(MH "Pamphlets")	1,997
S43	(MH "Television")	5,405
S44	(MH "Radio")	1,039
S45	(MH "Internet+")	79,516
S46	(MH "Social Media")	1,143
S47	S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46	132,409
S48	(MH "Information Seeking Behavior")	1,956
S49	(MH "Problem Solving")	5,175
S50	(MH "Problem-Based Learning")	1,530
S51	(MH "Professional Practice, Evidence-Based")	12,953
S52	S48 OR S49 OR S50 OR S51	21,405
S53	S47 AND S52	3,167
S54	(MH "Science") or (MH "Information Literacy") or (MH "Decision Making") or (MH "Critical Thinking") or (MH "Judgment") or (MH "Thinking")	29,790
S55	S53 OR S54	32,712
S56	TI (((health or information or mathematical or quantitative or science or scientific* or media) N2 (literacy or literate# or illiteracy or illiterate#))) OR AB (((health or information or mathematical or quantitative or science or scientific* or media) N2 (literacy or literate# or illiteracy or illiterate#)))	2,360

S57	TI (information N2 competen*) OR AB (information N2 competen*)	199
S58	TI numeracy OR AB numeracy	250
S59	TI ((scientific N2 (skill# or think* or reason*))) OR AB ((scientific N2 (skill# or think* or reason*)))	154
S60	TI ((critical N2 (think* or reason*))) OR AB ((critical N2 (think* or reason*)))	2,507
S61	TI evidence-based OR AB evidence-based	25,919
S62	S56 OR S57 OR S58 OR S59 OR S60 OR S61	31,034
S63	TI (((health or consumer or medical or scien*) N2 information)) OR AB (((health or consumer or medical or scien*) N2 information))	11,950
S64	TI risk information OR AB risk information	361
S65	TI (((health or medical or scien*) N2 (claim# or statement# or message#))) OR AB (((health or medical or scien*) N2 (claim# or statement# or message#)))	2,881
S66	TI (((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research))) OR AB (((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research)))	29,639
S67	TI ((media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#)) OR AB ((media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#))	127,550
S68	TI ((internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#)) OR AB ((internet* or world wide web or worldwide web or web site# or website# or web portal* or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#))	37,368
S69	S63 OR S64 OR S65 OR S66 OR S67 OR S68	200,514
S70	S62 AND S69	3,951

S71	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or consumer or medical or scien*) N2 information))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or consumer or medical or scien*) N2 information)))	1,015
S72	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 risk information)) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 risk information))	42
S73	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or medical or scien*) N2 (claim# or statement# or message#)))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or medical or scien*) N2 (claim# or statement# or message#))))	199
S74	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research)))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 ((health or medical or scien*) N2 (paper# or article# or report# or literature# or journal# or periodical# or research))))	1,934
S75	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N10 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine# or ((print* or written) N2 information) or pamphlet# or leaflet# or booklet# or brochure#)))	6,168
S76	TI ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 (internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#))) OR AB ((((critical* N0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) N7 (internet* or world wide web or worldwide web or web site# or website# or web portal# or blog* or web log* or bulletin board# or bulletinboard# or message board# or messageboard# or forum# or ehealth or e-health or electronic health or weblog* or moblog* or vlog* or video blog* or microblog* or wiki* or web page# or webpage# or chat room# or chatroom#)))	1,851

S77	S55 OR S70 OR S71 OR S72 OR S73 OR S74 OR S75 OR S76	46,114
S78	S1 AND S33 AND S77	377
S79	S14 AND S77	1,033
S80	S78 OR S79	1,349
S81	(MH "Experimental Studies+")	149,213
S82	(MH "Crossover Design")	8,333
S83	(MH "Repeated Measures")	34,049
S84	(MH "Quasi-Experimental Studies+")	7,163
S85	(MH "Multicenter Studies")	7,710
S86	(MH "Pilot Studies")	29,795
S87	(MH "Program Evaluation")	18,007
S88	TI ((randomis* or randomiz* or randomly or random allocat*)) OR AB ((randomis* or randomiz* or randomly or random allocat*))	94,139
S89	AB (group? and (random* or between* or control* or intervent*))	94,435
S90	TI ((multicenter or multi center or multicentre or multi centre)) OR TI trial OR AB trial	67,525
S91	TI ((intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) N5 (study or studies or design)) or (before N5 after) or (pre N5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect* or effectiveness or impact or time series or time point* or repeated measur*)) OR AB ((intervention* or controlled or control group or compare or comparison* or compared or ((prospectiv* or crossover) N5 (study or studies or design)) or (before N5 after) or (pre N5 post) or pretest or pre test or posttest or post test or quasiexperiment* or quasi experiment* or evaluat* or effect* or effectiveness or impact or time series or time point* or repeated measur*))	736,439
S92	S81 OR S82 OR S83 OR S84 OR S85 OR S86 OR S87 OR S88 OR S89 OR S90 OR S91	828,551
S93	S80 AND S92	884

Database: Science Citation Index Expanded og Social Sciences Citation Index
Date: 25.02.2014
Hits: 1,780

# 37	1,780	#36 AND #32
# 36	13,095,250	#35 OR #34 OR #33
# 35	12,547,480	TS=(intervention* or controlled or (control NEAR/0 Group) or compare or comparison* or compared or ((prospectiv* or crossover) NEAR/5 (study or studies or design)) or (before NEAR/5 after) or (pre NEAR/5 post) or pretest or (pre NEAR/0 test) or posttest or (post NEAR/0 test) or quasiexperiment* or (quasi NEAR/0 experiment*) or evaluat* or effect$ or effectiveness or impact or (time NEAR/0 series) or (time NEAR/0 point$) or (repeated NEAR/0 measur*))
# 34	955,121	TI=(multicenter or (multi NEAR/0 center) or multicentre or (multi NEAR/0 centre)) OR TS=trial
# 33	1,864,545	TS=(randomis* or randomiz* or randomly or random allocat*) OR TS=(Group$ and (random* or between* or control* or intervent*))
# 32	2,219	#31 OR #30
# 31	1,707	#29 AND #6
# 30	625	#29 AND #7 AND #1
# 29	81,735	#28 OR #27 OR #26 OR #25 OR #24 OR #23 OR #22
# 28	7,706	TS=(((critical* NEAR/0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) NEAR/7 (internet* or (world NEAR/0 wide NEAR/0 web) or (worldwide NEAR/0 web) or (web NEAR/0 site$) or website$ or (web NEAR/0 portal$) or blog* or (web NEAR/0 log*) or (bulletin NEAR/0 board$) or bulletinboard$ or (message NEAR/0 board$) or messageboard$ or forum$ or ehealth or e-health or (electronic NEAR/0 health) or weblog* or moblog* or vlog* or (video NEAR/0 blog*) or microblog* or wiki* or (web NEAR/0 page$) or webpage$ or (chat NEAR/0 room$) or chatroom$))
# 27	52,032	TS=(((critical* NEAR/0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) NEAR/7 (media* or television* or tv or radio or broadcast* or broadsides or news* or maga?ine$ or ((print* or written) NEAR/2 information) or pamphlet$ or leaflet$ or booklet$ or brochure$))
# 26	7,926	TS=(((critical* NEAR/0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) NEAR/7 ((health or medical or scien*) NEAR/2 (paper$ or article$ or report$ or literature$ or journal$ or periodical$ or research)))
# 25	595	TS=(((critical* NEAR/0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) NEAR/7 ((health or medical or scien*) NEAR/2 (claim$ or

		statement$ or message$)))
# 24	224	TS=(((critical* NEAR/0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) NEAR/7 (risk NEAR/0 information))
# 23	2,925	TS=((((critical* NEAR/0 read*) or apprais* or evaluat* or understand* or assess* or judg* or interpret*) NEAR/7 ((health or consumer or medical or scien*) NEAR/2 information)))
# 22	12,980	#21 AND #14
# 21	2,300,668	#20 OR #19 OR #18 OR #17 OR #16 OR #15
# 20	133,885	TS=((internet* or (world NEAR/0 wide NEAR/0 web) or (worldwide NEAR/0 web) or (web NEAR/0 site$) or website$ or (web NEAR/0 portal$) or blog* or (web NEAR/0 log*) or (bulletin NEAR/0 board$ or bulletinboard$ or (message NEAR/0 board$) or messageboard$ or forum$ or ehealth or e-health or (electronic NEAR/0 health) or weblog* or moblog* or vlog* or (video NEAR/0 blog*) or microblog* or wiki* or (web NEAR/0 page$) or webpage$ or (chat NEAR/0 room$) or chatroom$)))
# 19	2,030,828	TS=((media* or television* or tv or radio or broadcast* or broadsides or news* or magazine$ or ((print* or written) NEAR/2 information) or pamphlet$ or leaflet$ or booklet$ or brochure$))
# 18	129,437	TS= (((health or medical or scien*) NEAR/2 (paper$ or article$ or report$ or literature$ or journal$ or periodical$ or research)))
# 17	9,686	TS= (((health or medical or scien*) NEAR/2 (claim$ or statement$ or message$)))
# 16	1,703	TS= (risk N/0 information)
# 15	39,020	TS= (((health or consumer or medical or scien*) NEAR/2 information))
# 14	73,019	#13 OR #12 OR #11 OR #10 OR #9 OR #8
# 13	58,917	TS= (evidence-based)
# 12	4,329	TS= ((critical NEAR/2 (think* or reason*)))
# 11	1,748	TS= ((scientific NEAR/2 (skill$ or think* or reason*)))
# 10	1,198	TS= (numeracy)
# 9	617	TS=((information NEAR/2 competen*))
# 8	7,138	TS=(((health or information or mathematical or quantitative or science or scientific* or media) NEAR/2 (literacy or literate$ or illiteracy or illiterate$)))
# 7	2,469,772	TS=((educat* or train* or teach* or workshop$ or work-shop$ or seminar$ or course$ or curricul* or learn* or instruct* or self-instruct* or selfinstruct* or coach* or skill* or problem-based or pedagog* or class or classes or lesson$ or taught or module$))
# 6	135,466	#5 OR #4 OR #3 OR #2

# 5	24,396	(TS=((grade$ NEAR/2 6) or (grade$ NEAR/2 7) or (grade$ NEAR/2 8) or (grade$ NEAR/2 9) or (grade$ NEAR/2 10) or (grade$ NEAR/2 11) or (grade$ NEAR/2 12))) OR (TS=("year$ 6" or "year$ 7" or "year$ 8" or "year$ 9" or "year$ 10" or "year$ 11" or "year$ 12" or "year$ 13"))
# 4	11,957	TS=(((six or sixth or seven* or eight$ or nine or ninth or ten or tenth or eleven* or twelve or twelfth) NEAR/3 grade*))
# 3	53,861	TS=(((middle or secondary or high) NEAR/0 School$))
# 2	75,565	TS=((schoolchild* or school-child* or (school NEAR/2 student*) or pupil$))
# 1	239,767	TS=((age$ NEAR/2 (11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or eleven or twelve or thirteen or fourteen or fifteen or sixteen or seventeen or eighteen)))

Database: Sociological Abstracts (ProQuest)
Date: 25.02.2014
Hits: 445

((SU.EXACT("Longitudinal Studies") OR SU.EXACT("Research Design") OR SU.EXACT("Experiments") OR SU.EXACT("Empirical Methods") OR SU.EXACT("Quantitative Methods")) OR (randomise* OR randomize* OR randomly OR random allocate*) OR (group* AND (random* OR between* OR control* OR intervene*)) OR (interveneion* OR controlled OR control group OR compare OR comparison* OR compared OR prospective* OR crossover OR pretest OR "pre test" OR posttest OR post test OR quasiexperiment* OR quasi experiment* OR evaluate* OR effect* OR effectiveness OR impact OR time series OR time point* OR repeated measure*) OR ((before NEAR/5 after) OR ("pre" NEAR/5 post))) AND ((SU.EXACT("Secondary Education") OR SU.EXACT("Education") OR SU.EXACT("Special Education") OR SU.EXACT("Health Education") OR SU.EXACT("Sex Education") OR SU.EXACT("Courses") OR SU.EXACT("Curriculum") OR SU.EXACT("Teaching Methods") OR SU.EXACT("Teaching") OR SU.EXACT("Computer Assisted Instruction") OR SU.EXACT("Educational Programs") OR SU.EXACT("Literacy Programs") OR SU.EXACT("Educational Plans") OR SU.EXACT("Schools") OR SU.EXACT("High Schools") OR SU.EXACT("Junior High Schools") OR SU.EXACT("Secondary Schools") OR SU.EXACT("Junior High School Students") OR SU.EXACT("High School Students") OR SU.EXACT("Students")) AND (SU.EXACT("Literacy") OR SU.EXACT("Moral Judgment") OR SU.EXACT("Judgment") OR SU.EXACT("Participative Decision Making") OR SU.EXACT("Decision Making") OR SU.EXACT("Problem Solving") OR SU.EXACT("Thinking") OR SU.EXACT("Deduction") OR SU.EXACT("Inference") OR SU.EXACT("Reasoning") OR SU.EXACT("Induction") OR SU.EXACT("Evidence Based Practice")))

Database: Social Science Abstracts (ProQuest)
Date: 25.02.2014
Hits: 208

((SU.EXACT("Longitudinal Studies") OR SU.EXACT("Research Design") OR SU.EXACT("Experiments") OR SU.EXACT("Empirical Methods") OR SU.EXACT("Quantitative Methods")) OR (randomise* OR randomize* OR randomly OR random allocate*) OR (group* AND (random* OR between* OR control* OR intervene*)) OR (interveneion* OR controlled OR control group OR compare OR comparison* OR compared OR prospective* OR crossover OR pretest OR "pre test" OR posttest OR post test OR quasiexperiment* OR quasi experiment* OR evaluate* OR effect* OR effectiveness OR

impact OR time series OR time point* OR repeated measure*) OR ((before NEAR/5 after) OR ("pre" NEAR/5 post))) AND ((SU.EXACT("Secondary Education") OR SU.EXACT("Education") OR SU.EXACT("Special Education") OR SU.EXACT("Health Education") OR SU.EXACT("Sex Education") OR SU.EXACT("Courses") OR SU.EXACT("Curriculum") OR SU.EXACT("Teaching Methods") OR SU.EXACT("Teaching") OR SU.EXACT("Computer Assisted Instruction") OR SU.EXACT("Educational Programs") OR SU.EXACT("Literacy Programs") OR SU.EXACT("Educational Plans") OR SU.EXACT("Schools") OR SU.EXACT("High Schools") OR SU.EXACT("Junior High Schools") OR SU.EXACT("Secondary Schools") OR SU.EXACT("Junior High School Students") OR SU.EXACT("High School Students") OR SU.EXACT("Students")) AND (SU.EXACT("Literacy") OR SU.EXACT("Moral Judgment") OR SU.EXACT("Judgment") OR SU.EXACT("Participative Decision Making") OR SU.EXACT("Decision Making") OR SU.EXACT("Problem Solving") OR SU.EXACT("Thinking") OR SU.EXACT("Deduction") OR SU.EXACT("Inference") OR SU.EXACT("Reasoning") OR SU.EXACT("Induction") OR SU.EXACT("Evidence Based Practice")))

Database: OpenGrey.eu
Date: 01.06.2014
Hits: 124

(((health or information or mathematical or quantitative or science or scientific* or media) NEAR/2 (literacy or literate* or illiteracy or illiterate*)) OR (information NEAR/2 competen*) OR numeracy OR ((scientific OR critical) NEAR/2 (skill* or think* or reason*)))

Database: Social Care Online
Date: 16.02.2014
Hits: 40

- SubjectTerms:'"children"' including this term only
- OR SubjectTerms:'"school children"' including this term only
- OR SubjectTerms:'"adolescence"' including this term only
- OR SubjectTerms:'"young people"' including this term only
- OR SubjectTerms:'"adolescent boys"'
- OR SubjectTerms:'"adolescent girls"'
- OR SubjectTerms:'"young people"'

AND

- SubjectTerms:'"schools"' including narrower terms
- OR SubjectTerms:'"education"' including this term only
- OR SubjectTerms:'"health education"' including narrower terms
- OR SubjectTerms:'"sex education"' including this term only
- OR SubjectTerms:'"teaching"' including narrower terms

AND

- SubjectTerms:'"literacy"' including this term only
- OR SubjectTerms:'"numeracy"' including this term only
- OR SubjectTerms:'"critical thinking"' including this term only
- OR Title:'literac*'
- OR Abstract:'literac*'
- OR Title: 'critical thinking'
- OR Abstract: 'critical thinking'

Database: Social Science Research Network
Date: 02.06.2014
Hits: *See below*

“critical thinking”: 205 hits
“health literacy”: 21 hits
"scientific literacy": 10 hits
"information literacy": 58 hits
“media literacy”: 27 hits
“science literacy”: 5 hits

Database: Google Scholar
Date: 02.06.2014
Hits: 12,600
Commentary: Version 1, first 200 hits screened

((literacy) or (information competency) or (scientific thinking) or (scientific reasoning) or (critical thinking) or (critical reasoning) or numeracy) AND (education or school or student or teaching or training or course or class or curriculum or instruction or lesson)

Database: Google Scholar
Date: 02.06.2014
Hits: 32,000
Commentary: Version 2, first 200 hits screened

((health literacy) or (critical literacy) or (scientific literacy) or (critical thinking)) AND (education or school or student or teaching or training or class or curriculum or lesson or instruction)

Database: Google Scholar
Date: 02.06.2014
Hits: 20,700
Commentary: Version 3, first 200 hits screened

((health literacy) or (critical literacy) or (scientific literacy) or (critical thinking)) AND (education or school or student or teaching or training or class or curriculum or lesson or instruction) AND (health or body or medicine or human)

Database: clinicaltrials.gov
Date: 22.08.2014
Search: Advanced search: (Literacy OR literacies OR appraising OR appraisal OR critical thinking) AND Interventional Studies AND Child
Results: 91

Database: ICTRP Search Portal
Date: 22.08.2014
Search: Advanced search: (Literacy OR literacies OR appraising OR appraisal OR critical thinking): in title, condition or intervention. Search including all recruitment statuses.
Results: 235

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported in section*
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Objective
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Abstract**
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Background
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Objective and Eligibility criteria
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Methods
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Eligibility criteria
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Search strategy
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplementary material
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Study selection
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Data extraction
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Data extraction
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Assessment of risk of bias in included studies

PRISMA 2009 Checklist

Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Synthesis of results
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Synthesis of results

Section/topic	#	Checklist item	Reported in section*
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Risk of bias assessment
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Results and PRISMA flowchart
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Study characteristics in text and tables
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Risk of bias assessment and Supplementary material
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Effects of educational interventions
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A for meta analyses. Effects of educational interventions
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Risk of bias assessment and figure
Additional	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression	N/A

analysis		[see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Discussion
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Discussion
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Conclusion
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Funding

** When submitting final article manuscript sections will be replaced by page numbers.*
*** According to standards set by Centre for Evidence-based Practice, Faculty of Health and Social Sciences, Bergen University College for master's theses.*

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

Supplementary material 3

Table of excluded studies

Study	Reason for exclusion
Belland BR, Glazewski KD, Richardson JC. Problem-Based Learning and Argumentation: Testing a Scaffolding Framework to Support Middle School Students' Creation of Evidence-Based Arguments. Instructional Science: An International Journal of the Learning Sciences. 2011;39(5):667-94.	Not included intervention. Does not involve critical appraisal or such. Not included outcome. Does not report on critical appraisal or such.
Brand LG. Evaluating the Effects of Medical Explorers a Case Study Curriculum on Critical Thinking, Attitude toward Life Science, and Motivational Learning Strategies in Rural High School Students: ProQuest LLC; 2011.	Not included intervention. Does not involve critical appraisal or such. Not included outcome. Does not report on critical appraisal or such.
Chowning JT, Griswold JC, Kovarik DN, Collins LJ. Fostering Critical Thinking, Reasoning, and Argumentation Skills through Bioethics Education. PLoS One. 2012;7(5):8.	Not included health topic. Focus on genetics and ethics.
Froman RD, Owen SV. Can We Improve Science Literacy? 1994.	Not included study type. Does not include a control group. Not included outcome. Outcomes possibly reported for knowledge of science.
Gegner J, Mackay D, Mayer R. Computer-supported aids to making sense of scientific articles: cognitive, motivational, and attitudinal effects. ETR&D-Educ Tech Res Dev. 2009;57(1):79-97.	Not included outcome. Does not report on critical appraisal or such.
Keselman A, Kaufman DR, Kramer S, Patel VL. Fostering Conceptual Change and Critical Reasoning About HIV and AIDS. Journal of Research in Science Teaching. 2007;44(6):844-63.	Not included outcome. Focus on knowledge of facts.
Larson SC. The effects of academic literacy instruction on engagement and conceptual understanding of biology of ninth-grade students. Dissertation Abstracts International Section A: Humanities and Social Sciences. 2012;72(7-A):2349.	Not included intervention. Does not involve critical appraisal or such.
Marin LM, Halpern DF. Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. Thinking Skills and Creativity. 2011;6(1):1-13.	Not included intervention or outcome. Both the intervention and the outcomes measured involve several topics and are not limited to health related topics.
Mbajiorgu NM, Ali A. Relationship between STS approach, scientific literacy, and achievement in biology. Science Education. 2003;87(1):31-9.	Not included health topic. Focus on biology. Not included intervention. Does not involve appraisal or such.
Ross JA, Cousins J. Enhancing secondary school students' acquisition of correlational reasoning skills. Research in Science & Technological Education. 1993;11(2):191-205.	Not included health topic. Only half of the students are tested in health related scenarios.
Steckelberg A, Albrecht M, Kezle A, Kasper J, Muhlhauser I. Impact of numerical information on risk knowledge regarding human papillomavirus (HPV) vaccination among schoolgirls: a randomised controlled trial. Ger Med Sci. 2013;11:Doc15.	Not included intervention. Does not involve scientific literacy. Not included outcome. Focus on knowledge of facts.
Tsai P-Y, Chen S, Chang H-P, Chang W-H. Effects of Prompting Critical Reading of Science News on Seventh Graders' Cognitive Achievement. International Journal of Environmental and Science Education. 2013;8(1):85-107.	Not included health topic. Focus on genes and reproductivity.
Wilson CD, Taylor JA, Kowalski SM, Carlson J. The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. Journal of Research in Science Teaching. 2010;47(3):276-301.	Not included setting. Study laboratory-based summer school over two weeks.
Zeidler DL, Sadler TD, Applebaum S, Callahan BE. Advancing reflective judgment through Socioscientific Issues. Journal of Research in Science Teaching. 2009;46(1):74-101.	Not included study type. Classes randomised to intervention and control groups, but outcome only measured on some students from each class.
Zohar A. Fostering students' knowledge and argumentation skills through dilemmas in human genetics. Journal of Research in Science Teaching. 2002;39(1):35-62.	Not included health topic. Focus on genetics and ethics.

School-based educational intervention for enhancing adolescents' critical appraisal skills

Supplementary material 4

Population: Adolescents in grades 7 to 12
Setting: Lower and upper secondary schools in the US and Germany
Intervention: School-based educational interventions
Comparison: Other intervention or instruction as usual

Educational interventions that compare different teaching modalities

Outcomes	Comparison of groups		Relative effect (95%)	No. of participants (studies)	Quality of the evidence (GRADE)	Comment
	Abstracted instruction in causal reasoning	Situated instruction in causal reasoning				
Knowledge and skills relevant for critical appraisal Shortly following instruction						
Understanding causality Scale (test): 0 to 15 points (15 best) (converted to percentage scores, e.g score of 12 = 80%)	Mean percentage knowledge in the abstracted group was 69.9	Mean percentage knowledge score in the situated group was 14.68 higher (9.68 higher to 19.68 higher)		220 students (1 RCT) (33, 34)	⊕◯◯◯ VERY LOW[1, 2, 3]	Effect size (Cohen's d) was 0.7
Critical appraisal skills: Causal reasoning Direct testing, shortly following instruction	-	-	-	194 students (1 Non-RCT) (33, 34)	⊕◯◯◯ VERY LOW[1,2,3]	Only two students received a score of 3. 192 students scored 0. "..no significant differences existed among any of the instructional or transfer conditions"
Behaviour	Not reported		-	-	-	

Educational interventions compared to instruction as usual

Outcomes	Comparison of groups		Relative effect (95%)	No. of participants (studies)	Quality of the evidence (GRADE)	Comment
	Instruction as usual	Educational intervention in critical appraisal-related topics				
Knowledge and skills relevant for critical appraisal Shortly following instruction						
Understanding epidemiology Self-report. Scale: 5 to 25 points (25 best)	Mean posttest score was in the control group was17.94	Mean posttest score was (subgroups only[6]): 3.15* points higher in I-group 1 0.74 points higher in I-group 2 0.20 points higher in I-group 3 (CIs not reported)	-	512 students (1 cluster-RCT) (35)	⊕◯◯◯ VERY LOW[3, 4]	Posttest scores adjusted for pretest score, gender, ethnicity, first language, final grades, special education coded and unexcused absence (* = statistically significant difference, p<.0.5)
Direct skills (test). Scale: 0 to 11 points (11 best)	Mean posttest score was in the control group was 4.17	Mean posttest score was (subgroups only[6]): 0.71* points higher in I-group 1 0.80* points higher in I-group 2 0.26 points higher in I-group 3 (CIs not reported)				
Understanding EBM aspects Direct skills (test): Total score calculated as person parameters (Rasch model).	Mean posttest person parameters in the control group was 483 (SD 94)	Mean posttest person parameters in the intervention group was 114 parameters higher (85.65 to 142.35 higher).	-	255 students (1 Non-RCT) (38)	⊕◯◯◯ VERY LOW[3,5]	An increase in 100 person parameters regarded a relevant difference
Critical appraisal skills: Causal reasoning Direct testing shortly following instruction. Heterogeneous measurement scales	See comment	See comment	-	Study 1: 9 classes Study 2: 55 students (2 Non-RCT) (36, 37)	⊕◯◯◯ VERY LOW[5]	Study 1: Mean adjusted posttest score 2.27 in control classes; score was 1.34 "statistically higher" in intervention classes (no CIs, no p-value, max score possible 13) Study 2: Mean posttest score was 2.325 in control group; score was 1.26 higher in intervention group (p<0.01, max score possible 6)
Behaviour	Not reported		-	-	-	

[1]Sequence generation and allocation concealment not reported.
[2]Study conducted in one school only.
[3]Only one study
[4]Insufficient sequence generation, allocation concealment not reported, no adjustment for clustering effect and selective outcome reporting.
[5]No or insufficient adjustment for clustering effects and/or possible confounders, indirectness due to student population (special education) or education providers (researchers), and heterogeneity in measurement scales
[6]**Intervention group 1:** One more experienced teacher, 16 lessons (of 34 in total); **Intervention group 2:** Four teachers, 16-18 lessons; **Intervention group 3:** One teacher, 6-10 lessons

Supplementary material 5

Risk of bias assessments of primary studies *(ordered by study ID)*

Derry et al., 1998 (36)

Risk of bias

Bias	Authors' judgement	Support for judgement
Sequence generation	High	No random sequence generation due to study design: non-randomised controlled study, allocation based on self-selection (volunteers)
Allocation concealment	High	No allocation concealment, allocation based on self-selection (volunteers).
Baseline characteristics & outcome measurements comparable	Unclear	Group characteristics at baseline only described in text, no data. Although authors adjust for pretest score (personal communication with study authors), no details about differences/similarities between groups are provided.
Blinding of students and education providers	Low	Blinding of teachers and students not possible. Outcomes measured are objective (direct skills) and not likely to be influenced by lack of blinding. The post-test was administrated 10 days after the intervention was completed (p. 181). Although this might have allowed time for control students to prepare themselves for the test we consider this unlikely. We assess the risk of bias to be low.
Departures from intended interventions	High	Intervention and control teachers worked at the same school. It is possible that communication between teachers could have occurred. The authors report that the instruction unit was not implemented as ideally planned. Challenges with regard to student attitudes, discipline and physical arrangements of classrooms greater in one intervention classroom than the other (p. 191). We assess this domain to have high risk of bias.
Blinding of outcome assessment	Unclear	Interpretation of students' answers to test questions required judgment, and there was no information about whether outcome assessor was blinded.
Incomplete outcome data	High	Administrative error resulted in an entire class receiving the same pre- and posttest, the class was excluded from analysis.
Selective reporting	High	SDs or CIs for mean differences between intervention and control classrooms, the number of students in each group, or p-values are not reported.
Outcome measures reliable?	Unclear	No information on reliability measures for test instrument. We consider and internal consistency measures to have sufficiently high values
Outcome measures validated?	Unclear	Is one of the tests (John's Trial) less valid because it is conceptually more diverse from the instructional unit contents?
Other bias?	Unclear	No information about teacher characteristics and demographics in the participating intervention and control classes. Insufficient information about adjustment for confounding factors at student-level.
Overall assessment of risk of bias	High	

Hill 1998 / Hendricks 2001 (33, 34)

Risk of bias

Bias	Authors' judgement	Support for judgement
Sequence generation *Situated versus abstracted*	Unclear	No details about how the sequence was generated
Sequence generation *Transfer instruction*	High	Original randomization broken. Allocation of class periods based on researcher's decision.
Allocation concealment *Situated versus abstracted*	Unclear	Insufficient information to permit judgment of allocation concealment. "I acquired a list of both teachers' combined teachers and used it to randomly assign students to an instructional condition"
Allocation concealment *Transfer instruction*	High	Allocation of class periods based on researcher's decision.
Baseline characteristics & outcome measurements comparable	Unclear	Group characteristics at baseline are not reported, only characteristics for the whole population. No pretest measurement of outcomes.
Blinding of students and education providers	Low	Blinding of teachers and students not possible. Both groups received an active intervention. Outcomes measured are objective (direct skills) and not likely to be influenced by lack of blinding. We assess the risk of bias to be low.
Departures from intended interventions	Low	Teachers received training beforehand. Lesson plans were (partly) scripted and provided to teachers. All instruction in both groups was audiotaped to ensure fidelity of implementation. Non-instructional events and interruptions were equally distributed in each group (Hill, p. 122-3). Transfer instruction lesson was not audiotaped, but teachers received training and the single lesson provided was scripted. We assess risk of bias to be low.
Blinding of outcome assessment	Unclear	Interpretation of students' answers to some of the test questions required judgment, and there was no information about whether outcome assessors were blinded.
Incomplete outcome data	Low	Confusing information about how many students consented to participate (Hill, p. 133), the number reported is lower than the 220 students accounted for. Since the author specifies that 26 students are lost-to-follow-up for the transfer test (measured later than the first posttest) it is likely no loss-to-follow-up for the first posttest. Transfer test: 11% attrtion. Evenly distributed between groups.
Selective reporting	Low	No study protocol available, but no reason to suspect selective reporting.
Outcome measures reliable?	Low	Test instrument developed by authors. We consider test-retest reliability and internal consistency to have sufficiently high values.
Outcome measures validated?	High	Face validity only.
Other bias?	Unclear	The teachers' instructional behaviour was different (more animated teacher in situated group). The author does not provide information about what informed the selection of teachers and why the more animated teacher was allocated to the situated group. Transfer instruction: No information about correcting for matching, clustering effects or other confounding factors
Overall assessment of risk of bias *Situated versus abstracted*	Unclear	
Overall assessment of risk *Transfer instruction*	High	

Kaelin et al., 2007 (35)

Risk of bias

Bias	Authors' judgement	Support for judgement
Sequence generation	High	"[…] with the small number of volunteer teachers, we were not able to achieve a true randomization process and completely control for self-selection." (p. 25)
Allocation concealment	Unclear	Method of concealment is not described.
Baseline characteristics & outcome measurements comparable	Unclear	There are differences between values for baseline characteristics, but they are adjusted for in analysis. Although authors adjust for pretest scores, no details about differences/similarities between groups are provided. Given our judgment about sequence generation (high risk of bias), we consider the risk of bias to be unclear for this domain.
Blinding of students and education providers	Low	Blinding of teachers and students not possible. Although one of the outcomes (self-reported knowledge of epidemiology) is subjective, intervention and control groups were located at different schools. Informed consent from students was waived and they probably did not know they participated in a study. The outcome "knowledge of epidemiology" is an objective outcome (direct knowledge/skills) and not likely to be influenced by lack of blinding. We assess the risk of bias to be low for both outcomes.
Departures from intended interventions	High	Teachers were supposed to teach 30+ lessons. The number of lessons taught varied considerably between teachers (from 6 to 18 lessons). No measures were taken to validate instruction delivery (e.g. audiotaping). Although authors consider "intervention dosage" in their analyses we assess this domain to have high risk of bias.
Blinding of outcome assessment	Low	The two outcomes were assessed using a Likert-scale (self-reported knowledge) and multiple-choice test. It is not likely that unblinded outcome assessors make a difference.
Incomplete outcome data	High	Substantial attrition and unevenly distributed between groups: 34% in intervention group, and 17% in control group. These students may be different from students who participated.
Selective reporting	High	Only adjusted mean posttest scores for subgroups are provided, mean score for the total intervention group not reported. SDs or CIs not reported. Comparisons of changes from pre- to post-test only provided for the intervention sub-groups versus the non-randomised control group.
Outcome measures reliable?	Low	Test instrument based on previous scales, knowledge/skills test developed by authors. We consider test-retest reliability and internal consistency measures to have sufficiently high values.
Outcome measures validated?	High	Face validity only
Other bias?	High	Teacher characteristics and demographics (other than gender) not factored into analysis. Statistical analysis does not correct for clustering effects.
Overall assessment of risk of bias	High	

Leshowitz et al., 1993 (37)

Risk of bias

Bias	Authors' judgement	Support for judgement
Sequence generation	High	No random sequence generation due to study design: non-randomised controlled study, allocation based on self-selection (volunteers)
Allocation concealment	High	No allocation concealment, allocation based on self-selection (volunteers).
Baseline characteristics & outcome measurements comparable	High	Intervention: Special education students; Control: General education students. Although justified by study author "…to effect a more stringent test of the effectiveness of the instructional intervention than (…) had a comparable group of special education students served as a control group" risk of bias considered to be high due to differences in learning abilities between groups. The control group was not administered a pretest
Blinding of students and education providers	Unclear	Blinding of teachers and students not possible The teachers who provided the instruction were also participating. Could this have influenced how they acted towards students, did they try harder than other teachers would have done? There is no information about whether the lessons were scripted and/or observed. We assess risk of bias to be uncertain.
Departures from intended interventions	Unclear	Information about co-intervention or fidelity of implementation not provided. The control class was based in the same high school as some of the intervention group students and teachers, but contamination unlikely because intervention teachers were part of research team and intervention students were attending special education classes.
Blinding of outcome assessment	High	Interpretation of students' answers to test questions required judgment. Tests were scored independently by the two participating teachers. They were not blinded because they assisted some students in reading the test.
Incomplete outcome data	Unclear	No attrition in intervention group. We do not know if any students in the control class were absent because only student who were present the day of testing were included.
Selective reporting	Low	No study protocol available, but no reason to suspect selective reporting.
Outcome measures reliable?	Low	Although internal consistency was moderate (Cronbach's a = 0.51) the items in test tested independent dimensions of reasoning and inter-rater reliability was high (0.89).
Outcome measures validated?	Unclear	Validation not mentioned by authors.
Other bias?	High	Student teachers in intervention group had little full-time teaching experience. Teacher characteristics and demographics not collected and factored into analysis. Statistical analysis does not correct for clustering effects or confounding factors.
Overall assessment of risk of bias	High	

Steckelberg et al., 2009 (38)

Risk of bias

Bias	Authors' judgement	Support for judgement
Sequence generation	High	No random sequence generation due to study design: non-randomised controlled study, allocation based on researchers' decisions.
Allocation concealment	High	No allocation concealment, allocation based on researchers' decisions.
Baseline characteristics & outcome measurements comparable	High	The control group comprised Gymnasium students (personal communication with study author) and may therefore include more high-achieving students than the intervention group. No pretest measurement of outcomes.
Blinding of students and education providers	Unclear	Blinding of teachers and students not possible. Low motivation may have influenced test scores for intervention students in the second pilot because there were no more lessons before summer holidays for the parallel classes.
Departures from intended interventions	High	Some teachers of the intervention classes might also have taught control classes. Although researchers undertook instruction intervention class teachers were offered to be present. The curriculum changed from first to second pilot which limit the interpretation of results. We assess this domain to have high risk of bias.
Blinding of outcome assessment	Unclear	Interpretation of students' answers to some of the test questions required judgment, and there was no information about whether outcome assessor was blinded.
Incomplete outcome data	Unclear	Eight students (18%) lost-to-follow-up in intervention group. We do not know if any students in control classes were absent because only student who were present the day of testing were included.
Selective reporting	Low	No study protocol available, but no reason to suspect selective reporting.
Outcome measures reliable?	High	The instrument (Critical Health Competence Test) was still under development when study was conducted. According to the author the Rasch scalability was not yet achieved (39).
Outcome measures validated?	High	See comment for "Outcome measures reliable".
Other bias?	High	Teacher characteristics and demographics not collected and factored into analysis. Control for other confounding factors probably not done. Statistical analysis does not correct for clustering effects.
Overall assessment of risk of bias	High	