

4-2012

How well do young people deal with contradictory and unreliable information on line? What the PISA digital reading assessment tells us

Tom Lumley
ACER, Tom.Lumley@acer.edu.au

Juliette Mendelovits
ACER, mendelovits@acer.edu.au

Follow this and additional works at: <http://research.acer.edu.au/pisa>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#)

Recommended Citation

Lumley, Tom and Mendelovits, Juliette, "How well do young people deal with contradictory and unreliable information on line? What the PISA digital reading assessment tells us" (2012).
<http://research.acer.edu.au/pisa/3>

This Article is brought to you by the National and International Surveys at ACEReSearch. It has been accepted for inclusion in OECD Programme for International Student Assessment (PISA) by an authorized administrator of ACEReSearch. For more information, please contact repository@acer.edu.au.

How well do young people deal with contradictory and unreliable information on line? What the PISA digital reading assessment tells us¹

Tom Lumley & Juliette Mendelovits

Australian Council for Educational Research

With the advent of the Internet, infinite quantities of information have become available to almost everyone, and an ever-increasing proportion of reading, especially by younger people, takes place in digital environments. This entails new demands on readers. The traditional mechanisms in print publishing that exert some control over the reliability of information (Warschauer, 1999) are largely absent in the online environment. Operating successfully in the digital medium requires not only access to technology, but also the ability to integrate, evaluate and communicate information (Warschauer, 1999). Faced with large amounts of information and limited time, readers must continually make immediate evaluations of the usefulness of different sources, in terms not only of relevance but also of trustworthiness. Readers now need increasingly to make their own choices about which information to read, and which to trust.

There is sometimes an assumption that young people, as ‘digital natives’ (Prensky, 2001), are able to use online information effectively, including selecting and negotiating digital texts that are not only relevant for what they need, but also are likely to provide reliable information. This paper examines the question of how well young people are in fact able to recognise whether information is likely to be trustworthy. While some small-scale work has been done in this area (for example, Leu & Castek, 2006), this paper draws on data from the first large-scale international assessment of online reading, the Digital Reading Assessment (DRA) that was part of the Organisation for Economic Co-operation and Development Programme for International Student Assessment (OECD PISA)² in 2009.

When the reading framework that had been developed for PISA 2000 (OECD, 1999) was revised as the basis for the 2009 assessment, it was decided to include digital reading alongside print reading in the definition and elaboration of the domain, thereby recognising that reading proficiency in the 21st century must encompass reading in both media (OECD, 2010). With this inclusion came some important changes in the framework’s emphases, one of which relates to the increased onus on readers to evaluate text material, resulting from the deluge of unfiltered information that comes to us via the Internet. The shift in the framework’s emphasis is evidenced particularly in the revision of description of the three main reading aspects (or processes) on which the framework is built, from *reflect* (PISA 2000) to *reflect and evaluate* (PISA 2009). While acknowledging that evaluation of texts for accuracy, reliability and timeliness also takes place in the print medium, the PISA 2009 reading framework stresses that this aspect is even more crucial for proficient reading in the digital medium, and takes on somewhat

¹ Paper presented at the Annual Conference of the American Educational Research Association, Vancouver, April 2012

² For a detailed description of OECD PISA see Turner and Adams (2007).

different forms. The framework cites Rieh's (2002) identification of two distinct types of critical judgment that are called upon during online reading: predictive judgments and reflective judgments. *Predictive judgments* are made about which site to go to, based on relevance, authenticity and authority. *Reflective judgments* must be made once a site has been reached, about its authority, reliability, credibility and trustworthiness (OECD, 2010, p. 39).

Tasks used in the PISA 2009 digital reading assessment

A number of tasks were designed for PISA reading assessment to assess the level of 15-year-olds' proficiency in making both predictive and reflective judgments in the digital medium, where the capacity to deal with potentially unreliable or contradictory information is the focus of measurement. A field trial for the digital reading assessment was conducted in 2008 among the countries that intended to take the Main Survey (MS) assessment in 2009. A non-representative sample of about 200 students in each of the 19 countries was administered each field trial item. Subsequently, the MS DRA was administered to around 25,000 students, in 19 countries and 15 languages, in 2009. Some of the tasks were only used in the Field Trial (FT); others were used in both the FT and the MS.

EXAMPLE 1: ICE CREAM (USED IN FT ONLY)

In this task, students see a set of search results related to the search term 'ice cream'. The task requires them to make a *predictive judgement*, using only the information available in a set of search results, as the links do not have any content behind them. Ten links are available, from which students may select only one. Appendix 1 shows a screenshot of the first few search results.

This page shows search results for ice cream and similar foods from around the world. You want to know if ice cream can be part of a healthy diet. Which search result is most likely to give accurate advice? Click the button next to the link.

The result students are required to select to obtain credit is this:

- [National Food Information Centre: The Food Guidelines-Food Label Connection](#)
For example, the serving size for ice cream and similar confectioneries is 125 ml (a half cup), ... National Dietary Guidelines suggest you eat a diet providing 30 per cent or less of calories (joules) from ...
www.nfic.org/~dms/guidelines.html

Clues implying authority or trustworthiness are: 'National Food Information Centre'; 'National Dietary Guidelines'; and '.org'. The other results offer various levels of distraction.

EXAMPLE 2: SMELL (USED IN FT AND MS)

Go to the 'Food in the news' web page. Would this web page be a suitable source for you to refer to in a school science assignment about smell? Answer Yes or No and refer to the content of the 'Food in the news' web page to give a reason for your answer.

This task directs students to open a specified link and evaluate the content in terms of suitability for use in a school assignment. The page has a main article, 'The smell of pizza can change people's behaviour', and summarises 'a review of research on smell conducted by a leading European motoring organisation.' Other links on the page have an obviously commercial or sensationalist flavour: 'Entertainment', 'TV Guide', 'Your Say', and a series of other news stories such as 'The truth about soy sauce and cholesterol'. (See Appendix 1 for a screen shot of the text.)

In evaluating the web page, students could interpret 'suitability' in terms of either content (relevance, amount of information) or reliability (authority). PISA items are coded with a coding guide, which describes the features of each category, or code, that is given credit. Items may have only a single level of credit, or may have two (or more) levels, where responses indicating a higher level of reading ability are given full credit, and those indicating lower ability are given partial credit.

In constructing the coding guide for the FT, it was hypothesised that in this task, more able readers would be more likely to refer to reliability of the site than to content. Accordingly, responses that referred to the reliability and authority of the web page were initially given full credit, and those that referred to the relevance of the page were given partial credit. Sample student responses from the FT illustrate these typical ways of responding.

Coding Guide (summary) for Example 2: SMELL

Full Credit: Answers (or implies) No and gives a plausible supporting explanation, referring to the popular or sensational nature of the website content, or the popularisation of the issues by journalists.

Answers (or implies) Yes and indicates that the site would be helpful as a secondary source, leading to more reputable sources.

Partial Credit (FT): Answers (or implies) Yes and gives a plausible supporting explanation, referring to the article's sources of information or the level of detail provided.

The following are some samples of student responses to the *SMELL* question from the FT:

Full credit

Reliability/authority

- No because it has been written by a motoring company not a recognised scientific body. [*refers to lack of credibility of body commissioning the research*]

- No. I would not think this a reliable source to use in a science assignment. [*refers to reliability*]
- No, this would not be suitable for a science assignment. This would be good for a truck driver but not a science assignment. [*implies entertainment value of contents, lower standard of trustworthiness is required*]
- Yes, you could use this page in your science assessment, however, because it is not a well known site, it would be beneficial to use other, more well known websites to back up this information and research. [*refers to need for more reliable evidence*]

Partial Credit

Relevance

- YES; It shows us heaps of stats to include in an assignment [*refers to level of detail*]

No credit responses

Reliability/authority [responses offer no support for claims]

- Yes, as it is well researched and unbiased.
- Yes, the content is derived from a reliable source.

Relevance [responses focus on details only]

- Yes, it talks about how smell is the least understood of our senses among other points of interest. (no credit: focuses on details)
- No. This article talks about how the smell of pizza changes behaviour. It is not only about smell. (no credit: focuses on details)

The FT results showed that, in contrast with the test developers' hypothesis, students obtaining credit for their responses, those who talked about relevance (22%) had a higher mean ability than those who talked about reliability or authority, and there were much fewer of the latter (8%). As a result, for the MS, the two categories were collapsed into a single level of credit. In the MS, 25% of students obtained credit for this item.

EXAMPLE 3: LET'S SPEAK Q4 (USED IN THE FT ONLY)

This multiple-choice task presents students with a series of posts on an internet forum on the topic of speaking in public. It requires students to identify contradictory positions expressed in the posts.

Lauren writes, 'Even if you are very scared of speaking in public, there are things you can do to overcome your fear.' Which writer would be most likely to **disagree** with Lauren's statement?

The post that students need to select is this:

Julie: March 7 10:14

I think that the ability to speak in public depends on each person's personality. Some people seem completely incapable of public speaking. When they have to do it, their

hands shake and their voice trembles. Others, on the other hand, can discuss a subject fluently, in a way that makes the topic interesting for the audience. These people seem to be able to perform brilliantly, even if they have not had time to prepare! I'd say, there's no point in trying to change what you are.

The information needed is found in the last sentence of Julie's post: 'there's no point in trying to change what you are.'

EXAMPLE 4: LET'S SPEAK Q9 (USED IN THE FT ONLY)

This task requires students to make reflective judgements, articulating their criteria for evaluating the authority of information or sources of information. There are posts from six writers to the forum, while content for a seventh (Dr Nauckunaite) is available by clicking one of the links within the forum.

Look at Mischa's post for March 10. Click on 'Write a Reply' and write a reply to Mischa. In your reply, answer her question about which writer, in your opinion, knows the most about this issue. Give a reason for your answer. [Note: use the Back button to refer to the Forum page.]
Click 'Post Reply' to add your reply to the forum.

The coding guide use for the FT shows that two categories were hypothesised for quality of response: students referring to the professional status of either of two specific writers were expected to be of higher ability than those who referred only to content. The categories of response that were given credit are described in the coding guide.

Coding Guide (summary) for Example 4: LET'S SPEAK Q9

Full credit: Identifies Doctor Nauckunaite and/or Psychologist O.L. (explicitly or implicitly) AND refers to their professional status. May express scepticism about their professional status.

Partial credit: Identifies any of the four writers named by Mischa (Julie, Tobias, Psych OL or Dr. Nauckunaite) AND gives a reason that is consistent with the text, related to the cogency, practicality or logic of the text.

There was small difference in ability: those giving responses related to the professional status had a slightly higher ability than those referring only to the cogency, practicality or logic of the advice given.

PISA is reported in terms of 'levels', which indicated both the degree of difficulty of the task, and the degree of proficiency of the students. The described levels of proficiency are based on generalisations from the difficulty level of individual items administered in the assessment.

The PISA 2009 DRA was reported on four levels, Level 2 to Level 5, with Level 2 indicating the easiest items and the lowest level of digital reading proficiency, and Level 5 the most difficult items and the highest level of proficiency. Figure 1 shows the descriptions of the four DRA levels.

Figure 1. Summary descriptions for four levels of proficiency in digital reading

Level	Lower score limit	Percentage of students able to perform tasks at this level or above (OECD average)	Characteristics of tasks
5 or above	626	7.8%	Tasks at this level typically require the reader to locate, analyse and critically evaluate information, related to an unfamiliar context, in the presence of ambiguity. They require generating criteria to evaluate the text. Tasks may require navigation across multiple sites without explicit direction, and detailed interrogation of texts in a variety of formats.
4	553	30.3%	Tasks at this level may require the reader to evaluate information from several sources, navigating across several sites comprising texts in a variety of formats, and generating criteria for evaluation in relation to a familiar, personal or practical context. Other tasks at this level demand that the reader interpret complex information according to well-defined criteria in a scientific or technical context.
3	480	60.7%	Tasks at this level require that the reader integrate information, either by navigating across several sites to find well-defined target information, or by generating simple categories when the task is not explicitly stated. Where evaluation is called for, only the information that is most directly accessible or only part of the available information is required.
2	407	83.1%	Tasks at this level typically require the reader to locate and interpret information that is well-defined, usually relating to familiar contexts. They may require navigation across a limited number of sites and the application of web-based navigation tools such as drop-down menus, where explicit directions are provided or only low-level inference is called for. Tasks may require integrating information presented in different formats, recognising examples that fit clearly defined categories.

(OECD, 2011, p. 46)

Figure 1 shows that the capacity to ‘analyse and critically evaluate information’ – the kind of skill that has been discussed in this paper – appears only at Level 5.

Further, profiles of student performance at each of four empirically-derived proficiency levels include the following notes (OECD, 2011, pp. 49-50):

‘Students proficient at Level 5 ... are able to evaluate information from several web-based sources, assessing the credibility and utility of what they read using criteria that they have generated themselves.’ Only 8% of students, across OECD countries that participated in the DRA, performed at this level.

Students proficient at Level 4 ‘evaluate the authority and relevance of sources of information *when provided with support*’ (our emphasis). 30% of students in participating countries were proficient at Level 4 or above.

Students proficient at Level 3 have a more limited set of evaluative skills: ‘they evaluate information in terms of its *usefulness* for a specified purpose or in terms of *personal preference*.’ (our italics) – that is, they are capable of judging texts for relevance for their own purposes, but not in more disinterested terms, such as its authority or trustworthiness. A majority (61%) of 15-year-olds can exercise judgment at this level or above.

(OECD, 2011, pp. 49-50)

Drawing on data from the FT and the MS (where available), Table 1 shows the percentage correct and the proficiency level across the participating countries and economies for the four items shown above, and for two other items ('Secure 1' and 'Secure 2') with a similar focus that were fielded in the MS. The table indicates whether data are from the FT or the MS. Only those items used in the MS have been reliably related to PISA levels. Estimates for PISA levels are given for items used only in the FT. Table 1 includes reference to two tasks from the MS that are secure: the difficulty of these tasks and the focus of the tasks they present are described.

Table 1. Performance of students for sample items from PISA 2009

Item	Type of judgement	Percentage correct	Focus	PISA Level or Estimated Level ³	Data source
<i>Example 1:</i> Ice Cream	Predictive	31%	recognition of credibility	Estimated Level 4	FT 2008
Secure 1 Full credit	Predictive	27%	evaluation of credibility	Level 5 and above	MS 2009
Partial credit		34%	evaluation of relevance	Level 3	
Total		61%			
Secure 2 Full credit	Predictive	29%	evaluation of credibility	Level 4	MS 2009
Partial credit		34%	recognition of credibility	Level 3	
Total		63%			
<i>Example 2:</i> Smell	Reflective	25%	evaluation of credibility or content	Level 5 and above	MS 2009
<i>Example 3:</i> Let's Speak Q4	Reflective	49%	recognition of contradictory information	Estimated Level 3	FT 2008
<i>Example 4:</i> Let's Speak Q9 Full credit	Reflective	19%	evaluation of credibility	Estimated Level 5	FT 2008
Partial credit		25%	evaluation of content	Estimated Level 4	
Total		44%			

³ Estimates are based on a comparison between calibrations of tasks used in both the FT and the MS and tasks used only in the FT.

Results for the sample items presented here, and for the similar tasks which were included in the MS, showed, show that tasks demanding any kind of disinterested evaluation are relatively challenging for 15-year-olds, and those demanding the critical appraisal of texts for credibility or trustworthiness are particularly difficult. Appendix 2 shows student ability and difficulty of all items used in the FT. The item map shows that the items discussed in this paper (highlighted) all fall in the upper half of the distribution of difficulty, and (with the exception of LET'S SPEAK Q4, which focuses on contradictory information) above the ability level of the majority of students involved in the FT. The sample for the MS was different, and the scoring for these items was modified for the MS, and consequently the difficulty of the items was ultimately different.

Tasks requiring only recognition of contradictory information, or information that was potentially reliable or unreliable, tended to be answered successfully by between a third and half of the students. Once they were required to articulate an evaluation of the material's reliability, referring to the likely authority of the source, the proportion able to successfully respond fell to about a quarter of students. Only about 20% of students participating in the FT were able to explain why a particular source in the *LET'S SPEAK* unit was likely to have authority (because of professional training or academic status, although it is also possible that students are cynical about the value of those characteristics). This should concern teachers and policy makers, as it suggests that most 15-year-olds students do not know how to begin evaluating material they encounter on the internet. There is ample evidence that a majority of students consider it first in terms of relevance or interest, rather than looking at the reliability of its source.

Conclusion

The items used in the PISA 2009 DRA include a selection of tasks requiring students to make evaluations of both sources and content of pages available on the Internet, as well as to deal with contradictory information. This paper has shown some evidence that students are more able to evaluate the relevance of content they are presented with than its reliability. One task, requiring them to identify contradictory information, appeared to be relatively easier than tasks requiring evaluation.

There is an assumption that students are able to use online resources that are not only relevant to the tasks they are set, but also are likely to provide trustworthy information. In order to do this, students need to have criteria for evaluating information. This study provides empirical evidence that the great majority of 15-year-old students lack the ability to make basic evaluations about the credibility and trustworthiness of digital texts. Students need to be taught how to make these evaluations. This message is of critical importance to policy makers and teachers.

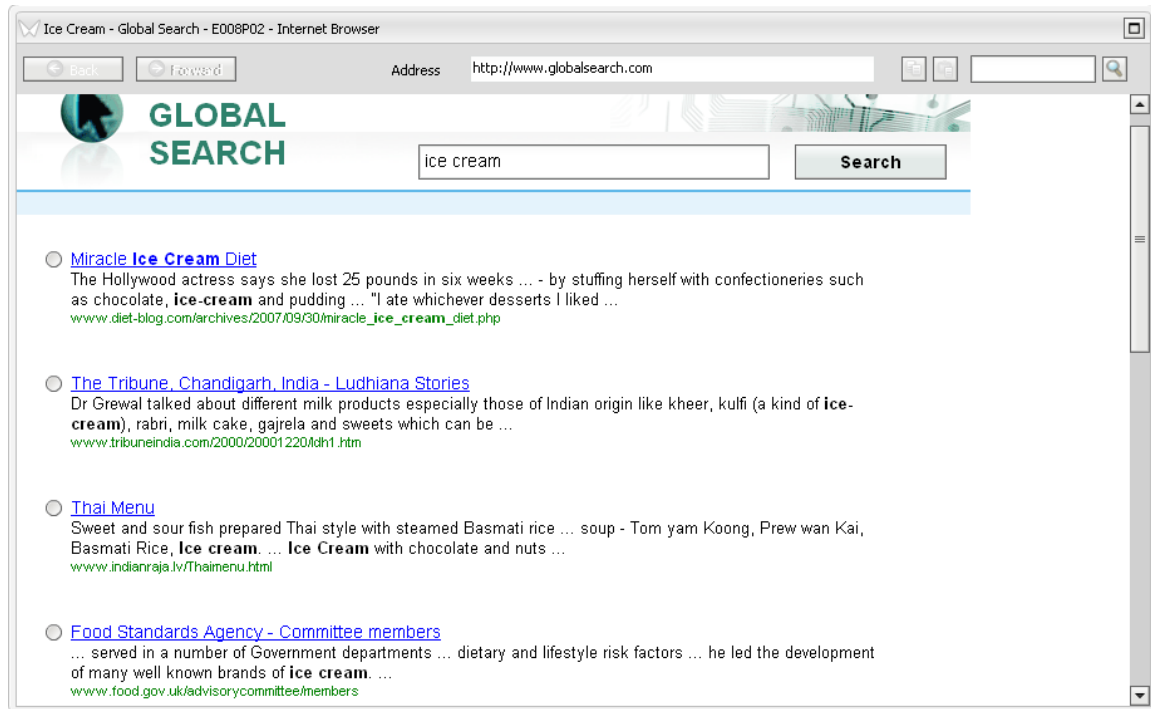
References

Leu, D., & Castek, J. (2006, April 9). *What skills and strategies are characteristic of accomplished adolescent users of the Internet?* Paper presented at the Annual

- Conference of the American Educational Research Association, San Francisco, CA.
- OECD. (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: OECD.
- OECD. (2010). *PISA 2009 Framework: Key Competencies in Reading, Mathematics and Science*. Paris: OECD.
- OECD. (2011). *PISA 2009 Results: Digital Technologies and Performance (Volume VI)*. Paris: OECD Publishing.
- Prensky, M. (2001). Digital Natives, Digital Immigrants. *On the Horizon*, 9(5): 1-6.
- Rieh, S. Y. (2002). Judgment of Information Quality and Cognitive Authority in the Web. *Journal of the American society for information science and technology*, 53(2), 145-161.
- Turner, R., & Adams, R. J. (2007). The programme for international student assessment: An overview. *Journal of Applied Measurement*, 8(3), 237-248.
- Warschauer, M. (1999). *Electronic literacies: Language culture and power in online education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix 1. Screenshots of selected items

Example2: ICE CREAM (information visible without scrolling: first four search results)



ICE CREAM: Task 1 [E008Q02]

This page shows search results for ice cream and similar foods from around the world. Which search result is most likely to provide a history of ice cream? Click the button next to the link.

Example 2: SMELL (information visible without scrolling)

The screenshot shows a web browser window with the address bar displaying 'www.whatsinthenews.com/articles/inn.asp?id=4381'. The page title is 'Smell - What's in the news? - E006P03 - Internet Browser'. The website header includes a compass icon and the text 'What's in the news? FOOD in the news'. The main content area features an article titled 'The smell of pizza can change people's behaviour' with a sub-headline: 'Researchers continue to show the power behind our sense of smell. Recent studies have found, among other things, that the smell of fast food wrappers or foods like pizza can cause road rage in drivers.' Below the article is a paragraph: 'This finding emerged in a review of research on smell conducted by a leading European motoring organisation. While some food smells can assist drivers in staying focused on the road, others can increase the possibility of driver error.' To the right of the article is a 'Related Stories' section with three links: 'Cheese makers to use electronic nose for market gain?', 'Horses can smell fear in humans', and 'The truth about soy sauce and cholesterol'. A navigation menu on the left includes links for 'Features', 'Entertainment', 'TV Guide', 'Shopping', 'Your Say', and 'Advertise with us'. A search bar at the top left contains the text 'Global Search What's in the news?'.

SMELL: Task 2 (E006Q05)

Go to the "Food in the news" web page. Would this web page be a suitable source for you to refer to in a school science assignment about smell? Answer Yes or No and refer to the content of the "Food in the news" web page to give a reason for your answer.

Note: The full set of PISA released DRA items is available from: <http://erasq.acer.edu.au/>
Retrieved 20 March 2012.

These credentials are needed in order to obtain access to the site:

User name: public

Password: access

Appendix 2. Map of Student ability and Item difficulty: FT 2008

Logits	Students	Items	Examples	Reflective/ Predictive	Cred/Relev Contradict*
3		31			
2		44			
		22			
	X	30			
	X	63			
	X	16 25	SMELL	Reflective	Cred/Relev
	XX	15			
	XXX	12 32			
	XXXX	26 53 68	ICE CREAM	Predictive	Cred
1	XXXX				
	XXXXXXXX	55 69	LET'S SPEAK Q9	Reflective	Cred/Relev
	XXXXXXXX	10 11 14 51			
	XXXXXXXX	5 9 49			
	XXXXXXXX	2 39 43 62			
	XXXXXXXXXX	45 50 57	{ SECURE 1 SECURE 2	Predictive Predictive	Cred/Relev Cred
	XXXXXXXX				
	XXXXXXXXXX	47			
	XXXXXXXXXXXX	42 66	LET'S SPEAK Q4	Reflective	Contradict
0	XXXXXXXX	18 38 64 67			
	XXXXXXXXXX	36 52 56 59			
	XXXXXXXXXX				
	XXXXXXXXXX	4 13 21 33 46 48 65			
	XXXXXXXXXX	40 58 61			
	XXXXXXXX	23 29 35 37			
	XXXXXXX	17 20 24 41 60			
	XXXXXXX	8			
	XXXX	28			
	XXXXXX	19 34			
-1	XXXX				
	XXX	27			
	XXX	3			
	XX	7			
	XX				
	X				
	X	6			
	X				
	X				
-2		54			
	X				
		1			

Each 'X' represents 25.0 cases
 *Cred = Credibility; Relev = Relevance; Contradict = Contradictory information