# Case Studies of Think Aloud Research in Science Education

John Oversby, Independent Researcher, [oversby61@gmail.com](mailto:oversby61@gmail.com)

## Abstract

Think Aloud research has received much attention in literature research but not in science education. This paper provides a social justice critique of Randomised Controlled Trials in terms of domination by the researcher perspective, and poses Think Aloud research as a kind of interview technique that hands much of the control to the participants, representing a form of increased power and influence for them. After some explanation of the Think Aloud Protocol in terms of the whole process, including analysis and interpretation, it considers four case studies where Think Aloud has been a significant feature of the research method. Three of these are in the domain of diagrams research, a central theme of the PALAVA teacher research group based in the UK. The other sought to explore pre-service teacher thinking as they answered examination questions for 16+ students. Finally, the paper concludes with aspects of Think Aloud that have been discovered that are pertinent to science education, showing the way for further development.

## Introduction

Education research, including science education research, has many different kinds, some of which are privileged by being preferentially funded. This paper includes, in part, a social justice lens to critique these kinds, especially in terms of the role and power of the participants. Of course, such research is highly influenced by research in other areas, not least that which takes place in the medical field. The latter is dominated by Randomised Controlled Trials but is also dominated by a positivist atmosphere, in which the participants are simply data sources in a research programme where the researchers are in control, experts with very strong power over the whole process. Even those in the control sample, where the intervention is presumed not to be as good from the start, are aliens in the exploration, hardly involved, and certainly given no authority. It is a system that has the greatest gap between the researchers and the participants while acknowledging that the system will have substantial impact on them, in some cases, life-determining. Modern medical research, in the Randomised Controlled Trails community, is one where the corporations involved hold the trump cards, always.

At one extreme, this corporate medical model, supported by the learned scientific societies as their champions, occupies part of the research space in education. The Evidence Based Teaching web site (<http://www.evidencebasedteaching.co.uk/>) claims ‘Like Evidence-Based Medicine, Evidence-Based Teaching uses methods which have been thoroughly tested and proven to be effective.’ They also treat teachers with some lack of respect for their skills in surveying the research literature: ‘While there is a lot of good evidence available, much of it is hidden in academic journals and books. 'The Evidence-Based Teacher’s Toolkit' presents this evidence in accessible and practical ways for teachers.’ This latter quote also fails to recognise efforts by education researchers to disseminate their work, seeing academic journals and books as the tools used by researchers to hide their findings. The Evidence-Based web site also notes: ‘U﻿﻿﻿sing the latest research, this video explains, in a jargon-free and accessible way, how the brain learns, and why some students find learning difficult.’ Implying that professional discourse is ‘jargon’ and that teachers need mediators to make this accessible. They are at pains to show that their work is based on ‘real lessons’, perhaps implying that education research is based on something else, maybe ‘unreal lessons’. Naturally, for teachers to access this mediated material, there is a substantial price tag to match. The Evidence-based Teachers’ Network (<http://ebtn.org.uk/about-ebt/what-is-ebt/>) sets out clearly what counts as good and poor evidence (the latter first given by the web site):

**Poor evidence**

Classroom experiments which have:

* a very short time scale
* few students involved
* no control group (to compare)
* lessons taught by the experimenters
* only one teacher involved

**Sources of good evidence**

Classroom experiments – conducted with control groups.  
Educational neuroscience – studying how the brain learns

Note here the caricature of extremes in the case of poor evidence, with the privileging of Randomised Controlled Trials, validated by some ‘educational neuroscience’. The rhetoric is very strong.

Finally, in this section, I should note the official UK government report (<https://www.gov.uk/government/publications/evidence-based-teaching-advancing-capability-and-capacity-for-enquiry-in-schools-interim-report>) which, rather unsurprisingly since it was commissioned by the Evidence-Based Teaching Network giving funding to academic colleagues to ensure that the message of such positivist education research is both adopted and disseminated.

Much is discussed in the science education research literature recently about the value of Randomised Control Trials (RCTs) (Chalmers *et al*, 1981), commonly used in analytical medical research. In England it has been proposed by the 2010-13 Coalition government as the means by which evidence-based research will form the basis for a new direction in research (proposal terminology, see <https://www.gov.uk/government/news/new-randomised-controlled-trials-will-drive-forward-evidence-based-research>). Acceptance of this approach, by the Wellcome Foundation for example. and in guidance from The National Foundation for Education Research (<http://www.nfer.ac.uk/publications/RCT01/RCT01.pdf>) is based mainly on analytical medical research and promoted by Ben Goldacre, an expert in such methods, (http://www.phc.ox.ac.uk/team/ben-goldacre). However, education is not medicine, and education research is not medical research. There are many challenges to simply adopting RCTs as a gold standard in education research:

1. RCTs accept that variables can easily be separated, distinguished and assessed. The analytical approach of much modern medicine supports such a view, while modern education adopts a more holistic attitude, intertwined with the fuzziness of much of human learning.
2. RCTs place the researchers as power privileged, treating participants simply as data sources. While such a power relationship can be ethical, it is suspect in terms of social justice. It adopts the principle that researchers are experts, and that participants are not, and therefore not able to make meaningful comment on the interventions proposed.
3. RCT surveys, questionnaires and interviews emanate from the researchers, and not at all from the participants, and are thus the product of the unequal power relationship.
4. Analysis of RCTs then continues to be in the hands of the researchers, often through statistical methods that further alienate the participants.
5. Interpretation of RCT analysis lies firmly in the hands of the RCT researchers, and rarely, if ever, in the hands of the participants.
6. RCT reports complete the separation of researchers and participants, often being written in expert and dense language, and published in inaccessible professional journals. It is rare for a version of a report to be written in lay language.
7. Finally, given their huge costs, RCTs can only be undertaken by fairly large, or large, research groups, making it nearly impossible for individual researchers, and especially schoolteachers, to conduct.

A major impact that led to the present strong focus on RCTs was the invitation of Michael Gove, the UK Secretary of State for Education to Ben Goldacre, a populiser of RCTs in medical research. He quotes on his web site http://www.badscience.net/2013/03/heres-my-paper-on-evidence-and-teaching-for-the-education-minister/

**Building evidence into education**

Dr Ben Goldacre will set out today how teachers in England have the chance to make teaching a truly evidence-based profession.

Education Secretary Michael Gove asked Dr Goldacre to examine the role of evidence in the education sector.

In a paper to be presented at Bethnal Green Academy, Dr Goldacre will say today that research into “which approaches work best” should be embedded as seamlessly as possible into everyday activity in education.

High-quality research into what works best can improve outcomes, benefitting pupils and increasing teachers’ independence. But Dr Goldacre’s recommendations go beyond simply running more “randomised trials”, or individual research projects. Drawing on comparisons between education and medicine, he said medicine had “leapt forward” by creating a simple infrastructure that supports evidence-based practice, making it easy and commonplace.

Dr Goldacre says that:

– research on what works best should be a routine part of life in education  
– teachers should be empowered to participate in research  
– myths about randomised trials in education should be addressed, removing barriers to research  
– the results of research should be disseminated more efficiently  
– resources on research should be available to teachers, enabling them to be critical and thoughtful consumers of evidence  
– barriers between teachers and researchers should be removed  
– teachers should be driving the research agenda, by identifying questions that need to be answered.

In some of the highest performing education jurisdictions, including Singapore, he explained: “it is almost impossible to rise up the career ladder of teaching, without also doing some work on research in education.”

Dr Goldacre said:

“This is not about telling teachers what to do. It is in fact quite the opposite. This is about empowering teachers to make independent, informed decisions about what works, by generating good quality evidence, and using it thoughtfully.”

“The gains here are potentially huge. Medicine has leapt forward with evidence-based practice. Teachers have the same opportunity to leap forwards and become a truly evidence-based profession. This is a huge prize, waiting to be claimed by teachers.”

The trail to preferring RCTs in education research is out in the open.

There are other criticisms of this recent move to prioritise funding of RCTs (see e.g. <http://blogs.lse.ac.uk/politicsandpolicy/theres-no-such-thing-as-a-free-rct-a-response-to-goldacre-and-gove/>) However, access to funding is a big influence, as the growth of the Education Endowment Foundation attests. For an example of its work on primary talking in science, see its web site (<https://educationendowmentfoundation.org.uk/evaluation/projects/thinking-doing-talking-science-effectiveness-trial/>) that is clearly based on an RCT. RCTs are best placed to search for analytical correlations between clear and unambiguous data. In practice there are some other problems that arise:

1. Control groups must be, at least, similar, if not identical. In practice this is assured on the basis of past assessments, such as examination or test results gained for a different purpose. It is then rare for the groups to be assured for age distribution similarity, although gender is usually one of the variables controlled. Social class distribution is difficult given the issue of confidentiality of personal home data, and proxy data, such as free school meals, or the number of books available at home, have to substitute, often without any further validation.
2. One composite variable is the teacher, the pedagogy used, and teacher beliefs and motivation. Much of the research focuses on the learner (student) outcomes, and takes little notice of difference in this composite variable. Also inherent in this variable is experience and expertise of the teacher.
3. Cause and effect are very difficult to extract from correlational evidence (see e.g. Beebee *et al*, 2012) We can only be sure of the correlations *in the circumstances and specific contexts of the data set*.

Some of us are interested in the changes going on in the mind as learning takes place. I am not convinced by the neuroscientists who claim that MRI imaging is the way forward (as though synapse bursts represent learning), nor am I convinced by much of the cognitive psychologists’ output, often conducted on undergraduate psychology students by postgraduate psychology students! The Think Aloud (TA) method which we have used in the PALAVA teacher researcher group for some substantial time now, is one we recognize that gets closer to exploring the changes that take place in learning. Perhaps even more significantly, the process of TA research seems to stimulate deep thinking about learning among teachers, and that is worth it whether or not TA is a faithful representation of learning. TA also distributes power in education research, as opposed to RCTs. The participants are not faceless individuals, data sources. Their thinking is personal collaboration with researchers, in an atmosphere of willing consent and commitment. It really I social justice in action. For other thoughts on these issues see Hammersley’s paper to the British Education Research Association in 2001, ‘Some Questions about Evidence-based Practice in Education’ (http://www.leeds.ac.uk/educol/documents/00001819.htm)

## Think Aloud Protocol

The TA method came from usability studies in 1982 but like most methods has developed considerably since (see, for example van Someren *et al,* 1994 for a thorough account*).* Essentially, the participant is asked to talk out their thoughts. The most common method is the Live or Concurrent TA, where the talk is recorded as the task is completed, with the emphasis on description rather than interpretation. It is also possible to use Retrospective TA. We have used the Live form. It could be said to be a form of an open interview, since the researcher’s role is simply to record (often audio) and to prompt for any more thoughts when the talk starts to flag. TA can often spring surprises, not least because control is much more given to the participant than in traditional interview techniques. One data analysis form is to transcribe the recording, with utterings, and then to subject the transcriptions to content analysis. This is very time consuming. In PALAVA, we have used a kind of Focus Group analysis, where the group members listen to the recordings iteratively, noting salient features and patterns, until *saturation* is reached, i.e. no new features or patterns are noticed. One advantage of this is that it also builds group cohesion and expertise. We claim that the process is less-directed than traditional interviewing but of course, prescribing the task is a form of researcher influence. It seems that researchers can give up some control of the research but not all! While the TA method is relatively easy to carry out, analysis is time consuming and can be challenging.

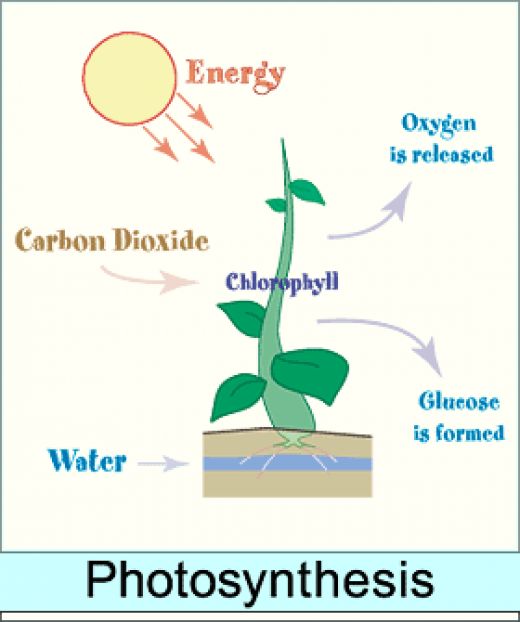
## Advantages of Thinking Aloud

Teachers are frequently orientated to how their students explain practical procedures, concepts, calculations and how to answer questions. The procedures of TA, that include simply listening, and prompts for more talk, are exactly those required for deep discussions. TA encourages teachers not to foreclose interesting discussions. The processes of setting a task for the thinking, and allowing for elaboration, mimics part of general class activity, though not so usually extended or focused on one or two individuals. In this sense, they have a degree of authenticity. For teachers who wish to carry out Action Research, using TA, the richness of the data provides valuable exploration of much that is of interest in the normal classroom. It is, of course, a qualitative method, but could sit alongside more quantitative methods as a form of triangulation to establish validity.

## Disadvantages of Thinking Aloud

By its very nature, TA is very time consuming, and demanding on the teacher remaining largely in the background. It is also very time-consuming in analysis and interpretation, as has been mentioned. It is quite possible that TA disrupts some normal thinking, since speech is so much slower than thinking. For younger participants, and with complex diagrams, the lack of appropriate vocabulary can also be a problem.

## Case Study 1: What do we see in scientific diagrams?

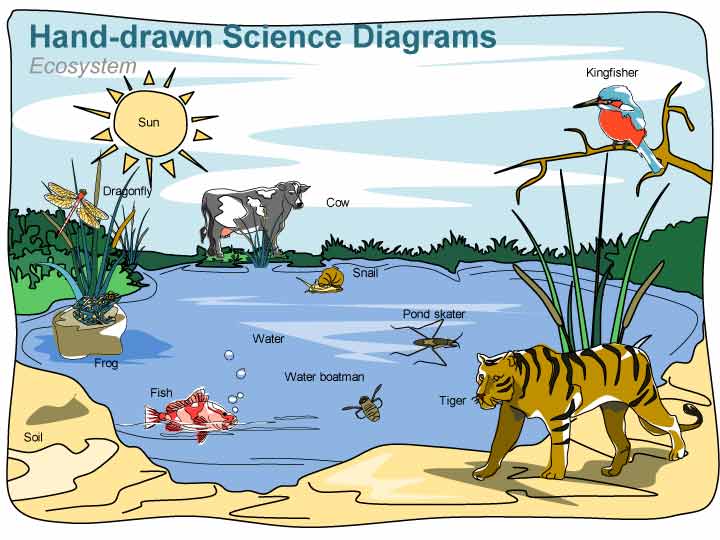


Typical of the scientific diagrams we used was one similar to that above. We used TA, with analysis and interpretation by the PALAVA group. We noted the following:

1. Generally, most diagrams used a convention where the ‘story’ moved from left to right, with inputs coming in from the left, and outputs going out on the right. As in the diagram above, we noted that the sun’s rays on photosynthesis diagrams were generally shown at the bottom right only, and generally moved only a little way down. This was also seen in depictions of candles where the rays were shown, although in this case, the rays were seen on every side.



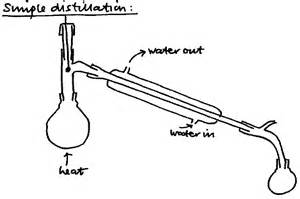
1. Generally, we used complex diagrams, as in the ecology system shown below. (Note the rays of the sun, which still go out only a short distance, then stop, but they do go in every direction.)



1. With all diagrams except for those showing cycles, the participants noticed sections starting, generally from top left, and moving to bottom right. In between there was some revisiting but the general pattern was strong.
2. Participants were frequently able to notice only a relatively few features of the diagrams, despite being prompted to tell us all that they saw. For example, using a stylised diagram of a flower, observations ranged from 12 to 76, the latter given by a person well used to observing diagrams.
3. When we gave a diagram with some obvious parts missing, such as a chemical formula, some participants told us that they saw these parts! From an associated experiment where we videoed (David Marsh did this) the participants’ eye movements (eye tracking) we found that the TA gave an average account of where the participant had looked over the past one or two seconds. The eyes moved very quickly. We found that they included, from memory, items that they expected to be present. This is not surprising since we know the sharp focus of the eyes is limited to the images that are on the fovea part of the retina. The rest of the image then sets the (fuzzy) context for the part being looked at. For diagrams that are complex, there is obviously much to be taken in, or assumed.
4. TA use with diagram ‘noticing’ gives us some insight into the brain’s operation, though we must always accept that this will be partial. However, we should not be too precious about this, since almost all forms of assessment provide only partial understanding of what is going on. The important thing is that we get some useful information.

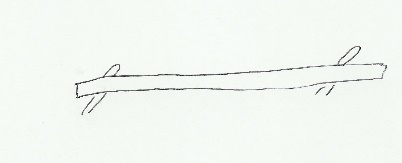
## Case Study 2: How does an individual construct a scientific diagram?

We asked participants (established teachers, and older students in schools) to construct a scientific diagram of equipment to separate two liquids by distillation.



We used direct observation by a group of teachers (members of the PALAVA teacher researcher group), and videoing over the shoulder, in conjunction with TA. The pattern we observed was that teachers of biology and physics, and school students, started with the condenser, then added the heating section, finally finishing up with the collecting flask. They explained that they drew the condenser first to make it ‘look right on the page’. The chemistry teachers generally drew the heating section first, then the condenser, then the collecting flask last, explaining as they went along the ‘story’ of the process in chronological order. We understood this difference through an appreciation of the strength of knowledge of the different teachers and students. With the advent of pre-drawn diagrams for PowerPoint presentations, such niceties of construction as a pedagogical instrument are being lost.

## Case Study 3: How do groups construct a scientific diagram?



We gave a ruler to three members of the group of teachers. As in the diagram above, we asked them to balance the rule on the index fingers. While keeping the ruler horizontal, we asked them to move the fingers together until they touched. They did this three times to get a feel for the activity, without any discussion of the science behind the activity. Then someone held one of the index fingers while they repeated the activity. Finally, we asked them to move the fingers apart from the mid-point. They were surprised by much of the activity (even the physicists), especially when they could only move one finger in the last part. Some even tried to wiggle the fingers to get the ruler ‘to do what it was supposed to do’! We then asked them to draw a diagram or diagrams, as a group, to explain to someone in the next room what to do (not to explain the physics). Working in a group prompted a kind of TA, although we noticed different levels of leadership and submission, rather than collaboration. In one case, a teacher separated from the other two and drew her own diagrams. We asked them to leave out text on the diagrams. We noted the following:

1. Working as a group was quite effective as a stimulant to TA, although it was limited by feelings of authority by different group members.
2. The participants were clear that they had considerable trouble in explaining the physics (they were all secondary science graduates).
3. They were unsure of what level of detail to include, often giving great detail on the fingers, or the graduations on the ruler. It seems that simplification is dependent on understanding what is going on.
4. We asked them to hold pencils in place of fingers, and to repeat the exercise. This time they drew the sharpened points of the pencils! Their discussions indicated that they still did not know how to explain the exercise.
5. Some groups used a story board method, with different diagrams to indicate the process, while some used arrows to indicate movement. This needs further exploration.
6. We came to the conclusion that group co-construction opens up different kinds of TA.

## Case Study 4: Exploration of Teacher Subject Matter Knowledge

With my colleague, Jane Fieldsend, we set about to find out whether pre-service teachers of science, all graduates of one science discipline or another, could answer correctly, typical English examination questions aimed at higher achievers in external GCSE examinations taken normally at 16+ years of age. The tests were given at the start and end of a one-year teacher preparation course. There were slightly more biologists. The post-test was at the same level with comparable questions, according to examination board information. The mean results for the pre-and post-tests were:

Pre-test (N = 88) Data are mean %

|  |  |  |  |
| --- | --- | --- | --- |
| Subject first degree | Biology questions | Chemistry questions | Physics questions |
| Biology | 61.3 | 51.3 | 38.8 |
| Chemistry | 58.3 | 75.6 | 53.6 |
| Physics | 56.6 | 61.2 | 71.9 |
| All students | 59.4 | 58.8 | 47.8 |

Post-test (N =74) Data are mean %

|  |  |  |  |
| --- | --- | --- | --- |
| Subject first degree | Biology questions | Chemistry questions | Physics questions |
| Biology | 75.3 | 74.0 | 57.8 |
| Chemistry | 83.7 | 93.1 | 88.0 |
| Physics | 77.2 | 90.0 | 95.2 |
| All students | 77.4 | 80.8 | 71.1 |

This rather stark data did puzzle us for some time. We had these concerns:

1. Given that the pre-service teachers were all science graduates across the disciplines, we wondered why they could not score more highly on questions aimed at 16+ year old students.
2. We wondered why biology graduates only scored three quarters in biology questions at the course end.
3. We wondered why biology graduates did so poorly in physics, despite having studied this top age 16+ previously.

We presented a sample of 12 student teachers, near the end of their course, with one physics, one chemistry, and one biology question, in a TA context. The recordings were discussed by three teacher educators. We were surprised by tow discoveries:

1. The students had used procedures they had memorized, and made no recourse to conceptual knowledge. We were testing recall of procedures, rather than problem-solving to answer a question.
2. When the student could not remember the procedure, or part of a procedure, they gave up completely. Only one student used a logical (not conceptual) approach.

The TA approach uncovered for us how the students answered these questions, whereas before we had only the bare evidence of the final outcome, the writing on the page. We decided we would be interested to know how the students taught examination technique in their practice schools, but did not have opportunity to carry out this research at that time.

## What have we learned from Think Aloud research methods?

1. TA method is appropriate for practising school teachers (and probably for Higher Education and pre-service teachers) to carry out qualitative research, including Action Research.
2. Data collection is relatively easy (as long as there is not too much prompting) but analysis can be both time-consuming and challenging. Great care should be taken to check validity with one or more colleagues.
3. TA is readily accessible to see the effects of an Action Research cycle, using the data from TA to modify teaching.
4. TA research can result in publishable material.
5. TA research can give *some* insight into brain action, which is clearly central to learning.
6. TA research places more control in the hands (or head) of the participants. This can make TA research more authentic.
7. TA research in science education is not yet prolific, compared with its use in reading, for example.
8. TA research can sometimes surprise us, as with the examples given above.

TA has caused us to think about social justice in education research.

|  |  |  |  |
| --- | --- | --- | --- |
| Method | RCTs |  | Think Aloud |
| Correlation or cause and effect | Correlation | 🡨---------------------------🡪 | Cause and effect |
| Funding required | Expensive | 🡨---------------------------🡪 | Cheap |
| HE v school-based | HE or research organisation | 🡨---------------------------🡪 | Every level, including school |
| Validity and reliability | Reliable but validity depends on the survey instrument used. Validity suspect if survey instrument only constructed by researcher | 🡨---------------------------🡪 |  |
| Role of participant v researcher | Privileged towards the researcher | 🡨---------------------------🡪 | Authentically collaborative between researcher and participant. |

## Implications for teaching and learning in science

TA is often used intuitively by teachers to explore a learner’s understanding of a problem. Here we show how this can be developed and converted into a valid and rigorous research activity.

## Further Research

TA can easily provoke ideas for further research, as I have indicated above. TA stands as a method that gives some control to the participants, which needs further exploration.

## References

Beebee H, Hitchcock & Menzies P (2012) The Oxford Book of Causation OUP Oxford

Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A (1981). "A method for assessing the quality of a randomized control trial". [*Controlled Clinical Trials*](https://en.wikipedia.org/wiki/Controlled_Clinical_Trials) **2** (1): 31–49. [*doi*](https://en.wikipedia.org/wiki/Digital_object_identifier):[*10.1016/0197-2456(81)90056-8*](https://dx.doi.org/10.1016%2F0197-2456%2881%2990056-8). [*PMID*](https://en.wikipedia.org/wiki/PubMed_Identifier) [*7261638*](https://www.ncbi.nlm.nih.gov/pubmed/7261638).

van Someren MW, Barnard YF & Sandberg JAC (1994) Think Aloud Method *A practical guide to modelling cognitive processes*, Academic Press, London, accessed at <http://echo.iat.sfu.ca/library/vanSomeren_94_think_aloud_method.pdf> June 2016

© John Oversby June 2016