Facing the evidence: Chernoff faces in evaluation.

Nova Reinfeld-Kirkman, *Ken Lountain*, Gill Westhorp.

Overview

This presentation addresses issues arising from a continuing collaboration between the authors in the evaluation of a South Australian Department of Education and Children's Services project "Maths for Learning Inclusion" (M4LI). The Program aims to systematically improve the engagement and learning outcomes of learners from low socio-economic backgrounds and Aboriginal learners through enhancing the capacity of primary teachers in the effective and inclusive teaching of maths. The focus of this presentation is the introduction of innovative data visualisation techniques (Chernoff faces) as part of this capacity building. The presentation will address the broad aims and evaluation of the M4LI project, provide a theoretical and practical context for the use of novel data visualisation techniques with this group, report on the initial evaluation of data visualisation process, and consider the implications of what has been learned.

The Program

Maths for Learning Inclusion is a South Australian program aimed at improving the teaching and learning of mathematics in selected primary schools serving low socio-economic communities. The Program was developed and is coordinated and implemented by the Learning Inclusion Team, led by Team Manager, Ken Lountain. The Team works within Curriculum Services in the South Australian Department of Education and Children's Services (DECS).

In program logic terms, the structure of Maths for Learning Inclusion suggests that:

- ⇨ employing a Coordinator to work with a cluster of schools, and
- ⇨ supporting the Coordinator and the staff with centrally-provided professional development programs will
- ⇨ lead to a local plan being developed, which will
- ⇨ underpin a range of program activities within the cluster, and that those activities will
- ⇨ lead to changes in teachers' attitudes, confidence and skills (these are program mechanisms at the teacher level), which will
- ⇨ underpin changes in teachers' teaching behaviours (which are both short term teacher-level outcomes and mechanisms for change in student learning behaviours), which will
- ⇨ lead to improved student engagement (which is a primary mechanism at the student level), which will
- ⇨ underpin improved student learning outcomes.

The initial phase of Maths for Learning Inclusion operated between 2005 and 2007. A second phase of the program has been underway since the beginning of 2009. Phase 2 has retained the main features of Phase 1 but has also included some key refinements drawn from what was learnt from the evaluation of the initial program.

The Evaluation

Education programs designed to improve teaching and learning in schools can be difficult to evaluate effectively. Programs may not be undertaken for long enough to generate significant data; insufficient resources may be allocated to evaluation; short lead times and the necessity for fast implementation may hinder the congruence of program design and evaluation strategies. Often, evaluation is 'tacked on' in the latter stages of programs or after their completion.

In designing and implementing the evaluation of the Maths for Learning Inclusion program the evaluation team actively set out to avoid these pitfalls and to produce a quality model of realist evaluation that can also be applied to other educational programs. The evaluation of the program is collaborative. It is supported by a Consultant, Gill Westhorp of Community Matters who specialises in Realist Evaluation.

A rigorous, ongoing evaluation was put in place from the beginning of the Program and provided the Program management team, the cluster coordinators, school leaders and teachers, and school communities with valuable ongoing information that helped shape the Program's direction.

Rather than operating on the periphery of the Program, the evaluation component of Maths for Learning Inclusion was an intrinsic element of the Program design, contributing to a culture of critical reflection and responsive action across the Learning Inclusion Team, the eight clusters, and 44 schools of over 200 teachers and nearly 4000 students.

Data sources for the evaluation comprised:

- Australian Council Of Educational Research Progressive Achievement in Mathematics tests (PATMaths) administered to students in Years 3, 4 and 5 in Terms 1 and 4; 2006 and 2007;[1]
- a 'Pre- and Post- Program Questionnaire' for teachers, addressing aspects of teacher knowledge, skill and attitude, their perceptions of leadership participation and support, Cluster Coordinators expertise, and access to professional development and peer support, again administered in Terms 1 and 4, 2006 and 2007;
- a shorter term-by-term questionnaire for teachers, collecting data about their participation in project activities (first year of program only);
- a term-by-term questionnaire for Cluster Coordinators, providing information about their activities, the adequacy of resources available to them, their approach to their role, and their perceptions of leadership participation and support (first year of program only);
- a Leadership questionnaire, seeking their perspectives in relation to their own participation in the project, the extent of leadership support for the project across the cluster, the adequacy of resources for the project, and relationships across the project;
- a focus group with leadership representatives from the eight clusters;
- a focus group in the first year, and paired interviews in the second year, with the eight Cluster Coordinators;
- 'Most Significant Change' stories written by teachers in all clusters.

---

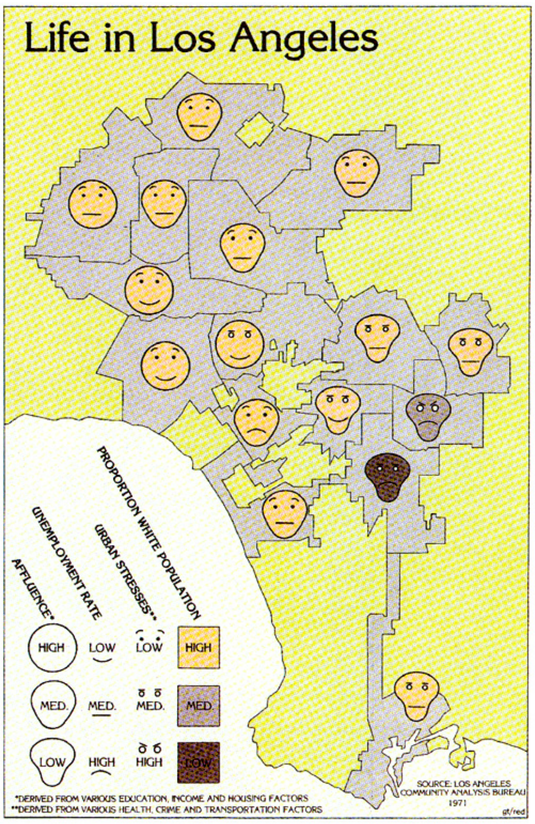
[1] The PATMaths is used in Maths for Learning Inclusion to provide information about student learning outcomes. It comprises a series of tests for different age levels, for which results can be converted to 'PATMaths scale scores'. This enable students' results to be compared over a period of years and thus allows comparison of student results over time.

These data sources and their analysis provided a unique and innovative balance of quantitative and qualitative data to assess the impact of Program strategies on improving teacher practice, student engagement and learning outcomes. However, only analysis of the first 2 data sources will be provided in the presentation.

Chernoff faces and M4LI

Toward the end of 2008, the M4LI team sought to develop strategies that would enable greater engagement with the data arising from the M4LI project. It was anticipated that this would be helpful both to those already in the project and those outside of it (i.e., potential participants, policy makers etc.). At this point, Nova Kirkman became involved in the M4LI team. Nova is an academic in the School of Medicine (Flinders University) with a responsibility for research and evaluation and a PhD Candidate in the School of Psychology (Flinders University). Nova's thesis is focussed on the determinants of responses to healthcare information. Her interest in data visualisation stems from the challenges inherent in presenting complex health-related statistical information to audiences of diverse academic backgrounds and differing levels of research training.

After initial discussions with Nova, the M4LI team decided to trial the use of Chernoff faces. Chernoff faces are graphical representations of multivariate data in the form of a human face (see Chernoff (1973). "The Use of Faces to Represent Points in K-Dimensional Space Graphically". Journal of the American Statistical Association 68 (342): 361–368). Each feature within the face (e.g., eyes, nose, ears) is used to represent a variable of interest. Changes in the facial feature (e.g., increasing in size, changes in the curvature of the mouth) reflect changes in the data which can be understood by using a a key or legend. For example, included below is the "Life in Los Angles" map by Turner (1979) (downloaded from…..)



The rationale for using Chernoff faces is based in sound principles of applied cognitive psychology. Humans are able to discriminate between seemingly limitless permutations of facial features (Chernoff, 1973), suggesting that the recognition of subtle variations in facial structures and/or expressions is an adaptive skill that is grounded in our evolutionary history (REF******) In fact, many non-human animals have been thought to display behaviours congruent with facial/ emotional recgonition (see, for example, Darwin ******), so these representations take advantage of 'deep' and automatic discriminative processing, enabling easier identification of trends within the data. (Chernoff teapots would probably not work!)

Because one is presented with 'faces' the emotional implications of the data presented are overt and undisguided. Turner (1979) said of his "Life in Los Angles" map that is was "probably one of the most interesting maps I've created because the expressions evoke an emotional association with the data." (cited in Zhou & Spinelli, 2004).

The emotional content of the faces has been suggested to add a "mnemonic advantage" (Chernoff, 1979) in so far as expressions are easily encoded and retrieved, allowing for greater retention and use of the data they represent. There is some evidence to suggest that Chernoff faces are more engaging (or simply less tedious) than other forms of data displays, such as complex tables (Scott, 1992).

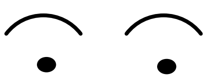
Additionally, the use of Chernoff faces as part of an evaluation opens up new avenues of research. It is well-established that apparently trivial decisions regarding formatting and presentation can affect the responses of participants when data is being *gathered*. Manipulations that increase the ease of visually processing a statement, such as heightening the contrast between text and background, also increase the likelihood that the statement will be rated as 'true' rather than 'false' ***(Schwartz, 2005)***. It therefore seems plausible that similar effects may be found when data is being *proffered*, rather than gathered. The role of data visualisation in evaluation programs appears to be under-researched, so there is little information regarding the effects (if any) of differences in data presentation in evaluation. A reliance on standard graphing techniques (Tufte, 1981) may exclude options that are potentially both more elegant and more useable.

The use of Chernoff faces in the M4LI program was first trialled at a professional learning session for school leaders with a focus on the sharing of project data, held in May 2009. Attendees were presented with graphs that showed for each year level and cluster:

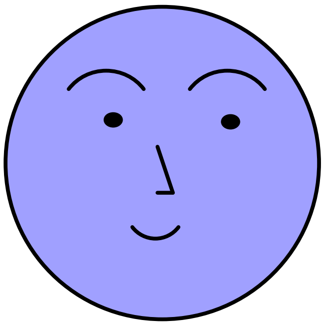
1. mean PATMaths scores for all students,
2. mean PATMaths scores for students identifying as Aboriginal or Torres Strait Islander.
3. The difference in mean PATMaths scores between school-card (disadvantaged students) and non-schoolcard holders was also calculated.

Participants were presented with a legend (example shown below) that was standardised by year level, so that scores reflected age appropriate standards. For each year level, the average scores for students within each group across all clusters was represented by the middle range of features. As students of Aboriginal and Torres Strait Islander background are a particular focus of this project, their scores were calculated separately using a different scale to that mapping general results. This was done in order to capture the variation in Aboriginal and Torres Strait Islander students' scores. As these student scores were, on average, lower than the general results using the same scale would have shown little in the range of ATSI students' scores.

LEGEND (Yr 5)

ATSI PAT Maths		PATMaths School Card vs. Non-School Card		PAT Maths (all)	
	Less than 35		Non-SC > 5+ points		Less than 45
	35-40		Non-SC > 0-5 points		45-50
	40+		SC > 0.1-5 pts		50+

Participants were then presented with 3 faces (one per year level) that showed their clusters' average scores. Faces were colour-coded by year level.



In the example to the left is one of the cluster's graph for Year 5 results. Using the legend above, we can see that Aboriginal and Torres Strait Islander students are performing, on average, better than the average Year 5 ATSI student across clusters. Non-school card holders are slightly better performing (0-5 pts) than school card holders, but this advantage is relatively small and within normal limits compared to the whole group. Likewise, mean PATMaths scores are equivalent to the average score across clusters (45-50pts).

Attendees were encouraged to discuss their results in comparison to the whole groups' norms, to identify points of similarity / dissimilarity, and to generate hypotheses concerning the cluster's current performance and the way in which it might be improved.

Evaluation of Chernoff faces

As part of the workshop evaluation, attendees were asked "Please list 3 adjectives or short phrases that describe your response to the Chernoff faces (e.g., engaging, confusing…)". These adjectives were coded as positive, neutral, or negative. "Confusing" was coded as "neutral", as because the valence of appraisal was often undefined. For example, "confusing until I learnt the key" suggests that the attendee was initially confused but came to understand the content. Coding "confusing" (and its permutations) neutrally also acknowledges that this is a cognitively demanding task, and is designed to be so. The table below shows a sample of adjectival feedback, and how it was coded for analysis.

Positive	Neutral	Negative
Multileveled, complex, interesting	Would like to know how to calculate info re own sites ATSI	Obtuse, unhelpful
Concise	2 participants drew Chernoff faces themselves…with no key	Not particularly helpful
Easily read (once key known)	Graphic	Usefulness????
Good idea to present data in diff form	Wide eyed, grimacing, snotty	Disturbing…
Fascinating, challenging, thoughtful	Confusing	Waste of time, added little, wonder why
	Unfamiliar	

For adjective 1, 5 persons did not provide any information. Of the remaining 41 attendees, 18 (43.9%) gave positive feedback, 15 (36.6%) provided neutral feedback, and 8 (19.5%) attendees' feedback was negative. As the list of adjectives continued, greater amounts of negative and incomplete feedback were found. Eight attendees (17.4%) did not provide feedback in adjective 2, however the greatest percentage of responses for adjective 2 were positive (44.7%, n = 17) or neutral (21.1%, n = 8), but more negative comments (34.2%, n =13) were found than in adjective 1. The third adjective had a large non-response rate 18 (39.4%) did not provide feedback for adjective 3, those that did generally reported positive (46.4%) or neutral (14.3%) responses, but again negative responses increased (39.3%). In total, 31% of responses were negative, 24% were neutral and 45% were positive.

Likert scales were also used to assess the extent of agreement (1= “strongly disagree”, 5 = “strongly agree”) with statements relating to the use of Chernoff faces in the workshop. For these specific questions mean responses were generally neutral (hovering around “3”), with high standard deviations. Interestingly, each question relating to Chernoff faces covered the complete range of possible responses (1-5). A table showing the descriptive statistics of these responses in shown below.

Item	Mean	Std Dev.	Range
The Chernoff faces helped me see comparisons in the data	2.36	1.20	1-5
Because I didn’t immediately understand the information in the Chernoff faces, I engaged to find out what they meant.	3.29	1.15	1-5
The Chernoff faces sometimes caused me to jump to inaccurate conclusions.	2.85	1.28	1-5

Paired t-tests were used to compare Likert-scale ratings of comparable questions in the usefulness of data presented in Chernoff faces or by traditional graphing methods. Teacher questionnaire information was presented in standard Excel bar graphs, and the utility of each set of information was assessed. Means of the item “Information from the Teacher Questionnaire will influence our program at school/ cluster level” were compared with means of the item ““Information from the PATMaths Tests will influence our program at school/ cluster level”. The difference between means was -.27 ($t(44) = -3.17$, $p<.005$), which is of clear statistical significance, but dubious practical significance. There was no statistically significant difference between the means of the perceived utility of breaking down the information in the Teacher Questionnaire or the PATMaths Tests by cluster. There is no way to evaluate how much higher (or lower) PATMaths ratings may have been if they had been presented in a traditional format, but mean evaluations were high (>4) and Standard deviations were relatively low(<1), indicating they probably would not have become much higher. At a minimum, if Chernoff faces were not particularly helpful at this stage of the evaluation they at least did no harm.

Overall, presentations at the workshop were rated highly. Mean ratings (SD) on a 5 point Likert scale were 4.1 (.76) for “I understand the information presented today”.

The Future…

There a number of reasons for only drawing tentative conclusions based on the information derived from this first evaluation of Chernoff faces in the M4LI project. Primarily, Chernoff faces were included in this first workshop in order to familiarise attendees with this form of graphing.

Chernoff faces are probably best used as a communication tool, rather than a reporting device. At this session, time constraints and competing priorities acted to attenuate the amount of discussion and engagement with the data represented in Chernoff face format. It is likely that having more time allotted to this task would have increased confidence and competence in reading the graphs, which would have then been translated into higher ratings of satisfaction. In addition, given the novelty of the format, it may have been strategic to put in place more structured activities than simple group discussion. For example, creating badges for the clusters (by year) giving one to each attendee and assigning tasks such as “find someone else with a different sort of nose” and then encouraging them to talk about their clusters and share ideas or information. A “speeddating" paradigm might also be useful (and fun). Ultimately, data is useful only in so far as it is used. Chernoff faces, and other forms of data visualization, offer the potential for creative, memorable engagement with complex multivariate data but the techniques for enabling this are yet to be refined.

Chernoff (1973). "The Use of Faces to Represent Points in K-Dimensional Space Graphically". Journal of the American Statistical Association 68 (342): 361–368

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. New York: John, Wiley & Sons

Tufte, E. (1983). *The Visual Display of Quantitative Information*. Cheshire, Conn: Graphics Press

Turner (1979) "Life in Los Angeles" downloaded from http://www.csun.edu/%7Ehfgeg005/eturner/gallery/lifeinla.GIF

Zhou, Y. and J. Spinelli (2004). Mapping Quality of Life with Chernoff Faces. Proceedings of Twenty-Fourth ESRI International User Conference. Down loaded from: http://proceedings.esri.com/library/userconf/educ04/papers/pap5000.pdf (24 July, 2009)