

---

Title	Using intervention-oriented evaluation to diagnose and correct students' persistent climate change misconceptions: A Singapore case study
Author(s)	Liberty Pascua and Chew-Hung Chang
Source	Evaluation and Program Planning, 52, 70-77
Published by	Elsevier

---

Copyright © 2015 Elsevier

This document is distributed under the Creative Commons Attribution-Non-Commercial-No Derivatives 4.0 International License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

This is the author's accepted manuscript (post-print) of a work that was accepted for publication in the following source:

Pascua, L., & Chang, C. -H. (2015). Using intervention-oriented evaluation to diagnose and correct students' persistent climate change misconceptions: A Singapore case study. Evaluation and Program Planning, 52, 70-77.

<http://dx.doi.org/10.1016/j.evalprogplan.2015.04.001>

**Notice:** Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.

## Using intervention-oriented evaluation to diagnose and correct students' persistent climate change misconceptions: A Singapore case study

Liberty Pascua\*, Chew-Hung Chang

National Institute of Education, Nanyang Technological University, NIE7-03-01, 1 Nanyang Walk, Singapore 637616, Singapore

### Abstract

The evaluation of classroom-based educational interventions is fraught with tensions, the most critical of which is choosing between focusing the inquiry on measuring the effects of treatment or in proximately utilizing the data to improve practice. This paper attempted to achieve both goals through the use of intervention-oriented evaluation of a professional development program intended to diagnose and correct students' misconceptions of climate change. Data was gathered, monitored and analyzed in three stages of a time-series design: the baseline, treatment and follow-up stages. The evaluation itself was the 'intervention' such that the data was allowed to 'contaminate' the treatment. This was achieved through giving the teacher unimpeded access to the collected information and to introduce midcourse corrections as she saw fit to her instruction. Results showed a significant development in students' conceptual understanding only after the teacher's decision to use direct and explicit refutation of misconceptions. Due to the accessibility of feedback, it was possible to locate specifically at which point in the process that the intervention was most effective. The efficacy of the intervention was then measured through comparing the scores across the three research stages. The inclusion of a comparison group to the design is recommended for future studies.

---

### 1. Introduction

Feedback is an invaluable resource in program planning, implementation and review. In the classroom, feedback offers a window for teachers to objectively assess the impact of an instruction on learning. Gathering feedback for teaching and learning program evaluation is usually implemented through formative assessment administered in the form of brief tests or short tasks, which provide actual data on the progress of learning and serve as prompts for immediate and timely refinement to the instructional approach (William & Thompson, 2009).

Collecting and incorporating responses for in situ teaching and learning intervention through traditional research methods pose several limitations. For instance, comparison groups may not always be available for the implementation of an experimental design. In addition, randomization is an issue that is often controversial. Indeed, the random assignment of students to control and experimental groups raises ethical questions and a practice that may not always be welcomed by teachers, school boards and parents (Check & Schutt, 2012; Monk, 1984). Conventional designs also disallow the contamination of treatment with the effects of treatment to ensure data fidelity, essentially blocking feedback from informing the program of intervention. This practice, however, does not optimize the data's latent capability to positively influence practice.

In this article, the authors explored the usefulness of intervention-oriented evaluation as a tool for classroom-based research cum professional development activity with teachers. The topic of interest was climate change, a subject whose tendency to be misunderstood by learners is well documented in the literature as well as in previous researches by the authors (Chang & Pascua, 2014). The intervention and evaluation designs were heavily adapted to the unique situations in the school setting, specifically the unavailability of a comparison group and the emphasis on self-reflection by the stakeholders as a key feature of the activity. Through a quasi-experiment, the evaluation aimed to consciously 'contaminate' the data, and successively incorporated corrections introduced to the overall analysis. With the aid of time-series design, the authors endeavored to expand discussion on alternative methods of formative evaluation whose flexibility allows for

\* Corresponding author. Tel.: +65 6790 3368; fax: +65 6896 9135.  
E-mail address: liberty.pascua@nie.edu.sg (L. Pascua).

non-conventional techniques that combine usefulness with reliable empirics without the need to sacrifice one for the other. Ultimately, the project intends to put forward recommendations for curriculum and pedagogical improvement for climate change education in Singapore.

### 1.1. Intervention-oriented evaluation

The conceptual and operational design of the activity followed Patton's (2008) concept of intervention-oriented evaluation, a sub-type of utilization-focused evaluation, which posits the evaluation process as a built-in instead of an add-on feature. With utility of evaluation as its core focus, the approach highlights the participation of intended users in the entire process of evaluation. In this study, the evaluation itself functioned as the 'intervention' such that the data was allowed to 'contaminate' the treatment. This was achieved through giving the teacher unimpeded access to the collected information and to introduce midcourse corrections as she saw fit to her instruction. Thus, the evaluation project was situational and indeed, even personal, in an environment of continuing cycle of reflection and innovation that allowed the data to show both the strengths and weaknesses of the intervention. Flowers (2010) commends this process for its inclusive approach which increases the sense of ownership and sustained involvement of participants.

### 1.2. Diagnosis and refutation of misconceptions

With the evaluation's operant design, it was necessary for a methodological approach that sanctions for repeated, unobtrusive testing and one that accommodates midcourse changes. Time-series was the favored approach to carry out the objectives of the research with data from the series made available for the teacher's perusal. Popularized by Gottman, McFall, and Barnett (1969), the time-series design is ideal for planned interventions when a control group is unavailable or when the parameters of an experimental design such as random assignment are not met. With only a few participants subjected to frequent and regular data collection, the method allows for the compilation of a wide variety of information on variables that interact in a research environment. Unlike the conventional pre-test and post-test approach, the time-series uses multiple baselines to establish pre-treatment conditions, which then becomes the basis of analysis on the degree of change (Check & Schutt, 2012).

Earlier papers in education research by Mayer (Farnsworth & Mayer, 1984; Mayer & Kozlow, 1980; Mayer & Lewis, 1979; Mayer & Rojas, 1982) delved into the development of the single-case time-series design in measuring student achievement and attitude. Its usefulness had also been recognized in special education studies (Horner et al., 2005; Marston, 1988), in measuring teaching effectiveness (Lin & Lawrenz, 1999), and self-regulated learning (Schmitz & Wiese, 2006). Time-series is preferred in longitudinal inquiries such as in the assessment of the effectiveness of school reforms (Bloom, 2003; May & Supovitz, 2006), the effect of financial aid to educational quality (Henry & Rubenstein, 2002) and the impact of lotteries to state expenditure for education (Moon, Stanley, & Shin, 2006). Ortega and Iberri-Shea (2005) posit that time-series "constitutes the single best formal strategy for investigating effects of instruction longitudinally" (p. 33) in their meta-analysis of approaches used for second-language acquisition. Despite its recognized usefulness, however, studies using the design remain few and far between when compared with the more popular pre-test and post-test approach.

### 1.3. Misconceptions and climate change

Climate change remains a misunderstood phenomenon despite its prominence in social and political discourses and its inclusion in various courses within environmental and geographic education. In fact, the literature attests to the prevalence of misconceptions in common knowledge about climate change causes, processes and impact across research contexts. This chronic issue is attributed to the prevalence of misconceptions in learners' mental schema. With a topic that is inherently complex and non-intuitive (McCaffrey & Buhr, 2008) the challenge in teaching climate change lies in on how to integrate misconceptions in mainstream pedagogical practices. Indeed, it had been reiterated that misconceptions bar efforts at improved climate literacy (Harrington, 2008; Dupigny-Giroux, 2010).

In context, this study in Singapore was conducted within a social backdrop of a government that has for the past decade dramatically increased its efforts to draw its citizenry's attention to climate change, its mitigation and related adaptation measures. Through the Singapore Green Plan (Ministry of Environment and Water Resources [MEWR], 2006), the state overtly communicated its acknowledgment of rising global temperature and its inherent threat to the island-nation's hard-earned development gains, its geography as well as resource vulnerabilities (Chang, 2014). A study by the National Climate Change Secretariat (Cheam, 2012) shows that many Singaporeans are keen to understand climate change in much more depth, and one suggestion put forward by the respondents was the inclusion of climate change in school curriculum. At present, climate change as a theme under environmental education is interspersed in Science and Social Studies, with Geography as the lead discipline in examining the world's climate system at the Secondary school level (Chang, 2014). In 2012, a joint syllabus revision exercise was conducted by the National Environment Agency and the Ministry of Education in which it was highlighted that environmental issues, particularly climate change, will feature more prominently in Secondary level curriculum (Cheam, 2012). Nonetheless, while the classroom figures as a key venue for environmental learning, young people in Singapore regard the media, particularly the Internet and television as the most ubiquitous sources of information on climate change (Chang, 2014; Tan, 2013).

Earlier research by the authors established baseline understanding on the lack of depth in the understanding of climate change by Singapore students (Chang & Pascua, 2014, in press). In addition, the misconceptions detected reflected to a great extent common confusions indicated in the international literature. Such include the confusion of the greenhouse effect with ozone layer depletion (Hansen, 2010; Lee, Lester, Ma, Lambert, & Jean-Baptiste, 2007; Österlind, 2005; Rebich & Gautier, 2005), the types and properties of greenhouse gases and the misunderstanding of the source, type and nature of the heat that is trapped (Boon, 2010; Gautier, Deutsch, & Rebich, 2006; Hansen, 2010; Mower, 2012; Österlind, 2005; Punter, Ochando-Pardo, & Garcia, 2011; Rebich & Gautier, 2005; Shepardson, Niyogi, Choi, & Charusombat, 2009). Confusions as to effects and management of the phenomenon are also prevalent. For instance, climate change is commonly perceived as related to certain environmental issues such as lead pollution, radioactive contamination and acid precipitation (Mower, 2012; Papadimitriou, 2004; Punter et al., 2011).

The reviewed literature is also denotative of the persistence of misconceptions prior to formal instruction and their tendency to obstruct proper learning (Duit & Treagust, 2003). Commonly referred to as misconceptions, these belief systems are formed from various sources of misinformation such as the media (Hansen, 2010) and weak instruction (Boon, 2010; Cordero, Todd, & Abellera, 2008; Lambert, Lindgren, & Bleicher, 2012). Misconceptions are not easy to detect (Chi & Roscoe, 2002) and efforts at implementing effective intervention have been documented to be particularly difficult and often unsuccessful (Limón, 2001). This is especially true when misconceptions are deeply rooted in the daily life

experiences of learners (Duit, 1999), such as through repeated exposure to unsubstantiated information on the Internet.

Studies that have reported on interventions to correct misconceptions concur on the seeming permanence of misconceptions. In their evaluation of the degree to which misconceptions are refuted in students' understanding of the greenhouse effect, Gautier et al. (2006) remarked on the resilience of misconceptions against instruction. Similarly, Lee et al.'s (2007) classroom intervention with elementary students yielded only partial improvement in conceptual knowledge with majority of misconceptions retained.

The tenacity of misconceptions is, to a degree, dependent on the nature of instruction. Often, instruction fails to consider for misconceptions and faulty linkages between propositions, resulting in a learner simply appending new information to an already existing body of misconceptions (Chi, 2008; Chi & Roscoe, 2002). In this light, this study aimed to simultaneously implement an intervention and evaluation aimed at determining pedagogical approaches that aid in the refutation of climate change misconceptions. The questions that guided the inquiry were:

- (1) Why are climate change misconceptions resistant to instruction?
- (2) What type of refutational instruction effectively corrects misconceptions?
- (3) To what extent does instruction for refutation result in partial and/or full revision of misconceptions?

## 2. Method

The research and evaluation was a collaborative professional learning activity with Geography teachers of a girls' school in Singapore. The intervention was conducted with a class of 31 Secondary 4 (Grade 9) students who were doing back classes for several units in Secondary 3 Geography, including the topic of climate change. Ms Leong (pseudonym), an experienced Geography teacher, taught the class. The key criterion for selecting this site was the teacher's and the department's keenness to participate in what they saw as a partnership to help students learn better. While Ms Leong had sole involvement in the activity, it was agreed that a review would be conducted with the department after its conclusion.

### 2.1. Data collection

Data was gathered, monitored and analyzed in three stages of a time-series design: the baseline, treatment and follow-up stages. There were 25 days of data collection with 8, 8 and 9 days for each of the stages, respectively. Individual class sessions for each of the stages lasted for an hour. For every meeting in the treatment stage, one-item questions were printed on individual pieces of paper that were passed to the students before a day's session ended. The responses were pooled daily to derive a mean score ( $N = 1$ ). Each mean score, referred to here as Class Score, was calculated through dividing the number of correct responses by the total number of students who were present in class. Essentially, the class score is a calculation of the collective level of difficulty experienced by the cohort vis-à-vis the lessons and testing as they progress. The formula used is similar to the calculation for the questionnaire's difficulty level in the pilot test (See Appendixes A.1 and A.4 for the formulae). No student received the same question twice. The Class Score was segmented into *General Content Understanding*, *Partial Understanding* and *Full Understanding* (See Section 2.4 for the scoring and analysis).

### 2.2. Intervention

Ms Leong learned of her students' misconceptions, as data from the daily testing was made available for her perusal. She occasionally consulted with the researchers to clarify her understanding of the topic and shared her reflections as to the progress of her students' learning. The authors did not prescribe any method or tool for teaching. Part of the inquiry was to ascertain how Ms Leong processed information; how, for example, she re-calibrated her approach based on the data and as a result of her interaction with students who exhibited misconceptions.

### 2.3. Identification of misconceptions

In order to ascertain if students have changed in their misconceptions, an instrument must be designed to diagnose what the areas of misconceptions are. Interviews with 27 students and a review of the literature informed the content of the diagnostic instrument used (Chang & Pascua, 2014, in press). Substantially, the bulk of the misconceptions identified referred to confusions on the causes and processes of climate change. The authors examined textbooks, syllabi and teaching guides to validate whether the items in the questionnaire were within the bounds of the grade level topic. It was confirmed that the items in the instrument did not deviate from the key features of the subject although the materials reviewed were generally 'silent' about misconceptions.

Treagust's (2012) two-tier diagnostic test was adopted for the questionnaire format. The test items' structure was composed of the Answer Tier and the Reason Tier in a multiple-choice format. The Answer Tier is a content question with 2–4 choices while the Reason Tier provides four or more possible explanations for the response in the Answer Tier (See Appendix B for examples. Email correspondence should be addressed to the article authors for the full set of questions).

The first draft of the instrument was subjected to face and content validity tests through consultation with a group of experts involving two Geography university lecturers, an education research specialist and a secondary school Geography teacher. The evaluators examined each item and gave suggestions for improvement. Overall, there was consensus on the format, content and relevance of the test items vis-à-vis the content and form of secondary school Geography.

The instrument was then used in a pilot test with 15 students of the same ability level with the sample class. Calculations for difficulty index, readability and reliability were conducted to ensure that the questionnaire fits the profiles of the target respondents. The test for difficulty measures the proportion of the sample that correctly chooses the correct answer combinations. It also detects which of the alternative choices are not effective distracters and have to be replaced (Nitko & Brookhart, 2011). The final 39-item version had a difficulty range of .13–.77; the questionnaire has items ranging from easy, moderate and difficult with most of the items converging on the moderate and difficult brackets. This was an expected result considering that the students were to formally learn the topic for the first time. Using the Flesch–Kincaid grade level test, it was ascertained that the instrument had a readability level of 8.4 confirming its suitability for the reading level of upper secondary school students (Grades 8–9). In addition, the Cronbach alpha score was calculated to measure how each item consistently relates to all other items in the questionnaire. With a .893 Cronbach alpha score, it was confirmed that the items have high internal consistency. The baseline score of .70 or moderate internal consistency is preferred in the social sciences (See Appendixes A.1–A.3 for formulae). Finally, the items were assigned arbitrary numbers in order to randomize its distribution in the whole test, a technique that is akin to the goal of random subject assignment (Monk, 1984).

#### 2.4. Scoring and analysis

Scoring the answers involved three ways. First, the overall Answer Tier score was computed as *General Content Understanding*. This was further segmented into a) the A-Tier score [AT] and b) AR Tiers score [ART]. The AT score refers to responses that identified the correct first tier answer but justified such with the wrong reason, indicative of *Partial Understanding*. ART scores, on the other hand, were responses that accurately identified the correct combination of answers in both tiers, thus a *Full Understanding*.

All the class scores were plotted in a line graph for visual examination of fluctuations and patterns across time. Statistical computations using Pearson  $r$  were employed to calculate for the correlation of scores for each of the three research stages. However, simple calculation of correlation does not account for autocorrelation of scores. Autocorrelation, or serial dependence, refers to how the value of one observation partly hinges on the value of one or more of the immediately preceding observations. It is a fact that autocorrelation is intrinsic in this study's design. As knowledge acquisition is an accrual process, what is learned in Day 1 adds on to new information learned in the succeeding days of data collection. Ignoring autocorrelation may lead one to overestimate the effect of treatment.

Borckardt's Simulation Modeling Analysis (SMA) for Time Series was used to complement the Pearson  $r$  scores. Apart from its ability to detect autocorrelation, the SMA allows for the analysis of short data series of less than 30 observations. Simulations of scores using multiple observations also further strengthened the decision to reject or accept the null hypothesis (see Borckardt et al., 2008 for the equation and calculation of SMA). Field notes, video recordings and an exit interview with the teacher were useful tools in the validation of statistical data.

### 3. Results and discussion

The three stages of the research were defined by different content topics discussed. Class sessions for the treatment stage commenced with the topic of Climate Change on the 9th day of data collection. Prior to this, the students were learning about the Elements of Weather and Climate in the baseline stage. Finally, the treatment stage ended upon the introduction of a new topic, Tourism, for the follow-up stage on the 17th day.

Scores in a time-series are expected to increase noticeably at the onset of treatment. Upon withdrawal or discontinuation of an intervention, the scores are projected to plateau or return to baseline level. In this study however, the expected sharp increase in understanding did not happen within the first three days into the treatment stage. This is despite the fact that instruction was completed during this period on the causes and processes of climate change that comprise the bulk of students' misconceptions (Fig. 1).

A closer look at the teacher-chosen pedagogy employed gives insight as to the nature of misconceptions and their resilience. On the first day of treatment, the teacher asked two students to illustrate and explain to the class their understanding of the climate change process, including all elements that they deem related to the phenomenon. Misconceptions that surfaced were then addressed at the end of the activity with the teacher explaining to the class verbally the main elements that interact in the enhancement of the greenhouse effect. The computed class score for the day registered to be even lower than the scores in the tail-end of the baseline stage.

On the second day, the teacher taught the concepts of heat trapping vis-à-vis the role of greenhouse gases. In her lecture, she interspersed misconceptions identified in the previous meeting. Such included the falsity of the belief that the ozone layer is the culprit to climate change, the types and concentrations of greenhouse gases, and the process of heat-trapping. She also explained why chlorofluorocarbons (CFCs) are not the main greenhouse gases that cause global warming while reiterating the differences between infrared and ultraviolet radiation. On this day, fossil fuels were discussed for the first time. The teacher did not refute misconceptions on the third day of treatment. Scores for the two sessions were climbing rather gradually, indicative that most were unable to fully grasp the concepts taught.

The fourth day was pivotal to both the learning process and the evaluation. On this day, the teacher was supposed to transition to another topic. However, upon review of the three-day test results and in consultation with the researchers, Ms Leong decided to alter her approach. She devoted most of the class time to the explicit refutation of misconceptions. As a rehash, she listed all known misconceptions on the board and asked for the students' opinion of whether a concept is related to climate change. She compartmentalized concepts into different categories (e.g., greenhouse effect

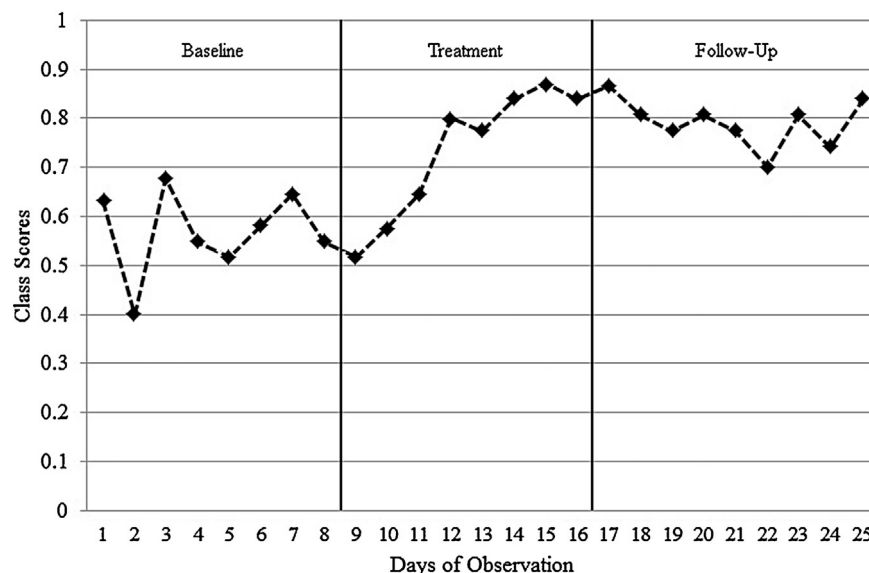


Fig. 1. General content understanding.

Table 1  
Simulated modeling analysis (SMA) for general content understanding.

Test statistics	Treatment (A), baseline (B) (8, 8) <sup>b</sup>	Follow up (A), treatment (B) (9, 8)	Follow up (A), baseline (B) (9, 8)
Pearson <i>r</i>	$r = .61, p = .0063$	$r = .30, p = .1246$	$r = .86, p = .0001$
Autocorrelation est. for Phase A <sup>a</sup>	$.70, p = .0004$	$-.15, p = .4561$	$-.15, p = .4561$
Autocorrelation est. for Phase B	$-.67, p = .0181$	$.70, p = .0004$	$-.67, p = .0181$
Simulation modeling (10,000 observations)	$p = .1041$	$p = .5442$	$p = .0001$

<sup>a</sup> Phases A and B refers to the two stages in comparison. For example, between treatment and baseline, the former is Phase A and the latter is Phase B, as indicated in parenthesis in the column headers in the tables.

<sup>b</sup> Correspondingly, the numbers in parenthesis refer to the days of data collection for each respective stages.

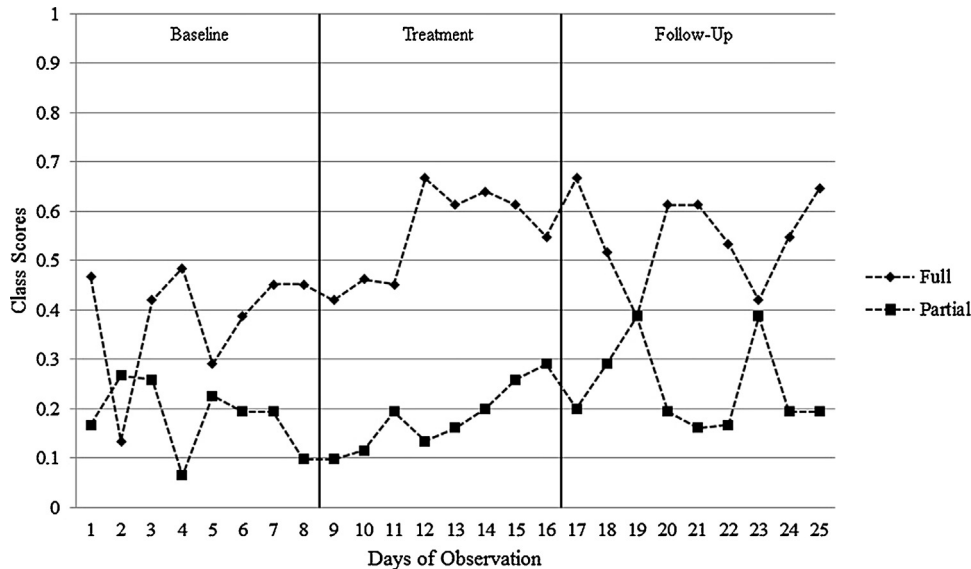


Fig. 2. Full and partial understanding.

Table 2  
Simulated modeling analysis (SMA) for full understanding.

Test statistics	Treatment (A), baseline (B) (8, 8)	Follow up (A), treatment (B) (9, 8)	Follow up (A), baseline (B) (9, 8)
Pearson <i>r</i>	$r = .64, p = .0038$	$r = -.014, p = .4749$	$r = .65, p = .0023$
Autocorrelation est. for Phase A	$.49, p = .0226$	$-.06, p = .5885$	$-.06, p = .5885$
Autocorrelation est. for Phase B	$-.36, p = .1843$	$.49, p = .0226$	$-.46, p = .1189$
Simulation modeling (10,000 observations)	$p = .0285$	$p = .9798$	$p = .0006$

vs. ozone depletion; shortwave vs. longwave radiation; excessive solar penetration vs. heat-trapping, water vapor and carbon dioxide vs. CFCs, etc). She allowed the students to openly challenge assumptions and beliefs about the topic. The collective score for the day improved dramatically. The teacher discussed the Copenhagen Accord and the Kyoto Protocol on the 5th Day (Day 13). Two other instances of refutation were performed on the 6th (Day 14) and 7th (Day 15) days in the treatment stage.

Statistical calculations for Pearson *r* confirm that successful refutation took place in the treatment stage only ( $r = .93, p = .01$ ). There was no evidence of a learning pattern in the baseline stage ( $r = .10, p = .81$ ) and a negative correlation, although insignificant, was observed in the follow-up stage ( $r = -.32, p = .41$ ). Further, SMA calculations that compared scores between the stages of the research corroborate that misconceptions were refuted significantly in the treatment stage ( $r = .61, p = .0063$ ) while momentum effect or delayed learning was not evident in the follow-up stage ( $r = .30, p = .1246$ ). When counting for forgetting as a factor, there was strong evidence that refutational instruction generally improved students' *General Content Understanding* ( $r = .86, p = .0001$ ). Autocorrelation was evident in the treatment stage ( $p = .0004$ ). With serial dependence taken into account, the simulation of scores showed that the null hypothesis could be rejected ( $p = .0001$ ) meaning there was improved understanding post-instruction, but also cautions on whether such is due to the efficacy of the intervention ( $p = .1041$ ) (Table 1).

Could improvement in *General Content Understanding* mean an increase in *Partial* or *Full Understanding*? An investigation of the line graphs showed an upward pattern for both scores (Fig. 2). Further, a segmentation of the *General Content Understanding* score indicated that there was stronger evidence of revision of misconceptions over plain assimilation of information. There was a moderately strong and significant correlation between treatment and baseline scores ( $r = .64, p = .0038$ ) for *Full Understanding* (Table 2). On the other hand, there was no significant correlation observed for *Partial Understanding* scores (Table 3). Overall, there was a moderately high and significant improvement in *Full understanding* ( $r = .65, p = .0023$ ) and the null hypothesis is rejected ( $p = .0006$ ) (Table 2).

Finally, an analysis of video recordings, field notes and interview with the teacher confirmed that the use of implicit refutation through contesting misconceptions only as they surfaced in

Table 3  
Simulated modeling analysis (SMA) for partial understanding.

Test statistics	Treatment (A), baseline (B) (8, 8)	Follow up (A), treatment (B) (9, 8)	Follow up (A), baseline (B) (9, 8)
-----------------	------------------------------------	-------------------------------------	------------------------------------

students' responses or as an appended element to a pre-determined content of instruction was less effective. Indeed, the students' reception to the nature and content of instruction on the day the teacher used direct and explicit refutation, as if they were learning the concepts for the first time, is testament as to how tenacious misconceptions are to the teachers' expositions on the misconceptions. The questions asked and reactions by the students clearly reflected the unraveling of errors in their mental schema. The following two excerpts from the videos demonstrate this:

“So, ozone depletion is not related to climate change? But the hole allows in more ultraviolet light.”

“Ahh, carbon dioxide! CFCs only a bit. CFCs mainly for ozone.”

While the intervention has clearly helped the students come to terms with the unraveling of errors in their mental schema that expedited the refutation of misconceptions, the time-series design also granted the authors and the teacher an opportunity to pinpoint exactly where the program was most effective. In turn, it allowed the authors to evaluate the change in the instruction that is most effective for this intervention.

#### 4. Conclusion

The evaluation of classroom-based educational interventions is fraught with tensions, the most critical of which is choosing between focusing the inquiry on measuring the effects of treatment or in proximately utilizing the data to improve practice. This study attempted to achieve both goals through the use of intervention-oriented evaluation of a professional development program intended to diagnose and correct students' misconceptions of climate change. Aided by the time-series design, data was gathered, monitored and analyzed in three research stages: the baseline, treatment and follow-up. The teacher as participant to the evaluation was given unimpeded access to the data. Thus, feedback was integral to the data pool; the data itself was at the heart of the intervention that was employed in the treatment stage, effectively affecting the treatment.

Through her interactions with the research team, the teacher was made fully aware of misconceptions held by her students even prior to her initial discussion of the topic of climate change. She employed refutational instruction from the very start of the treatment phase in what will be dubbed here as *reactive* refutation, or the identification and correction of incorrect answers as they surfaced. The results of the quasi-experiment highlighted how the students responded to her pedagogical approach. Initially, there were no observable and positive movements in the class scores. There were two factors at play that define such resistance to refutation. First, this can be attributed to the deeply ingrained nature of misconceptions. New and correct concepts introduced in class were unsuccessful in replacing wrong concepts, rather, the former were simply appended to the students' pre-existing understanding of the issue. Second, the resilience of misconceptions is also linked with the teacher's instructional approach of reactive refutation. While this type of pedagogy recognizes the presence of misconceptions, it was not directed so much at refutation of identified errors in the students' conceptions. There were no confirmatory measures, for instance, to take into account whether misconceptions were successfully refuted or not. The act of refutation, in this sense, was not the end-goal of the teacher's pedagogy.

The contrast in pedagogical approaches became evident when the teacher employed an approach referred to here as *pro-active* refutation on the 4th, 6th and 7th days of treatment. This type of refutation was direct, explicit, iterative and centred on identifying and correcting misconceptions in the open. The dramatic increase of class scores was indicative of this shift in understanding. Indeed, resilient misconceptions require multiple revisions using explicit rather than implicit refutation (Chi, 2008; Chi & Roscoe, 2002). The results also showed that direct refutation facilitated the development of full rather than partial understanding.

This collaborative activity underscores the 'soft' strengths of intervention-oriented evaluation. The involvement of the teacher, indeed of the whole Geography department, in the evaluation process moderated the impersonal nature of the research. Since the data was not withheld, ownership of the process was established from the very beginning and the teacher was empowered to determine for herself the strengths and weaknesses of her approach. In the end, the results of the evaluation were not prescriptive; rather, they were regarded as products of self-reflection and guided discovery of what works and what does not by all parties.

Certain recommendations are put forward based on the results of the quasi-experiment. First, there is a need to enhance teachers' knowledge of persistent misconceptions. While an educator's content knowledge on the subject may be adequate, there is generally lack of cognizance of the presence of misconceptions and the detrimental effects this brings about in the learning process. Second, while refutational instruction is efficacious in correcting misconceptions, the data also shows that a reactive, one-time revision is not enough to re-organize deep-seated beliefs, emphasizing the need to revisit and confront misconceptions time and again. A caveat here is timely. While the teacher was able to use refute misconceptions, it is also a fact that she would not have been able to do so without the research team's input of their persistence. The teacher may have developed a clearer understanding of the topic but she was blind to the fact that her first attempt at refutation did not fully untangle stubborn misconceptions. In hindsight, the teacher shared that there was initially no conscious decision on her part to change her style of teaching for the research. She assumed her students already grasped the concepts accurately on the first day of instruction.

Third, pedagogical content knowledge should be developed in teacher's instruction repertoire with due emphasis on the incorporation of misconceptions in addition to having a pro-active approach to conscious refutation of misconceptions. Lee et al. (2007) also suggested for students' misconceptions and learning difficulties incorporated into curriculum materials and teacher professional development. A focus on the related but individual concepts is first necessary to scaffold students toward more complex understanding of problems and concepts.

There are limitations to the research and issues with data collection that must be specified. First, the manner in which iterative refutation was implemented was not experimentally controlled. However, it must be reiterated that the sessions were back classes for an unfinished Geography syllabus. The primary objective of class meetings was to cover all of the topics in the shortest time possible with the most efficient of means. Indeed, time is a foremost concern with the national exams looming in the horizon. Could the topic

have been taught better? Without a comparison group, this question cannot be answered at this point. A stronger time-series design could make use of a comparison group. The lack of materials that equip teachers with the knowledge of misconceptions was also cited, as the current teaching materials and textbook are generally silent to the problem. It is therefore recommended that refutational texts and pedagogical training be available to complement text- books and teaching guides.

Finally, this research confirms the complexities intrinsic to climate change education and to the project of building functional literacy on climate change in general. Competing voices and polarizing views from climate advocates, alarmists and deniers alike have by far, held down knowledge building on the phenomenon. This research took the discourse of the issue from the global stage to the local level, and presented an intimate documentation of how this climate phenomenon is being (mis)understood on the ground, in the classrooms, by the next generation of planetary citizens.

##### 5. Lessons learned

An intervention-oriented evaluation complements the unique characteristics of classroom research with all its resources and limitations. Two components of the design used, the time-series and the two-tier diagnostic test, were keys to the successful implementation of the activity. The time-series design's strength lies in the frequency of data collection. Due to the method's responsiveness to feedback, it was possible to introduce design alterations and locate specifically at which point in the process when instruction effectively refuted misconceptions. A pre/post- test method would have only allowed a one-dimensional perspective through a single baseline score and post-treatment assessment. Through data gathered in a series of multiple baselines, observations and testing, there were enough instances of triangulation to confidently isolate key points when the intervention had an effect, and pinpoint to the strategy that delivered results.

The two-tier exam format complemented the time-series design. On the surface, students may provide correct propositional knowledge, but providing a second tier to test whether reasons attached to such propositions are correct or not may prove that there are, in fact, loopholes to their understanding. In addition, adopting the single-case approach ( $N = 1$ ) afforded for an unobtrusive, intensive, long-term data gathering and iterative approach to the diagnosis and refutation of misconceptions. The combination of gathering time-series data with a retrospective analysis of the teacher's pedagogy provides insight into the persistence of number of students taking the answer

While a strong case is posited in this case study on the capacity of an operant design aided by time-series data, the authors do not encourage for the results to be interpreted for the whole population. Indeed, there are other factors that are equally important for a holistic investigation such as gender, school culture, and teachers' content and pedagogical content knowledge among others. Nonetheless, this inquiry stipulates a deep and exhaustive assessment of the learning process and provides insight in ways to assist, engage, equip, and work together with teachers in designing responsive and intelligent formative evaluation methodologies that fit into their individual classroom cultures. A replication of the research with explicit refutation as the stand- alone intervention is hereby recommended.

In considering the evaluation of any education program, the approach adopted in this article becomes a useful tool for curriculum makers in designing an intervention program that works. This will enable curriculum designers to identify and collect useful elements, which can then be kept and improved in future iterations of the program. In adopting a Tyler (1949, 2013) approach to curriculum design in which it is carried out in the cycle of design, implementation, evaluation and review, the methodology advanced in this article is both highly effective and useful, to this end.

##### Acknowledgements

This paper refers to data from the research project "The hole in the sky is causing global warming?: restructuring prior knowledge in climate change education through conceptual change" (RS 5/12 CCH) funded by the Research Support for Senior Academic Administrator (RS-SAA) Grant, National Institute of Education (NIE). The views expressed in this paper are the authors' and do not necessarily represent the views of NIE.

##### Appendix A. Formulae

###### A.1. Difficulty level (pilot test)

$$p = \frac{\text{number of students choosing the correct answer}}{\text{number of students taking the test}}$$

###### A.2. Flesch–Kincaid grade level test (pilot test)

$$(.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$$

where ASL = average sentence length (the number of words divided by the number of sentences),  
ASW = average number of syllables per word (the number of syllables divided by the number of words).

###### A.3. Cronbach alpha test for reliability (pilot test)

$$\alpha = \frac{N \times \bar{c}}{\bar{v} + (N - 1) \times \bar{c}}$$

where  $N$  = number of items,  $\bar{c}$  = average inter-item covariance among the items,  $\bar{v}$  = average variance.

###### A.4. Class score (quasi-experiment)

$$\text{Class Score} = \frac{\text{number of students choosing the correct answer}}{\text{number of students taking the answer}}$$

##### Appendix B. Sample questions

1. In general, greenhouse gases destroy the ozone.
  - A. True.
  - B. False.
2. The reason for my answer is:
  - a. Trapped heat depletes the ozone layer.
  - b. Apart from chlorofluorocarbons, other greenhouse gases do not damage the ozone.
  - c. Carbon dioxide is the gas mainly responsible for the depletion of the ozone layer.
  - d. Others: \_\_\_\_\_
3. Sea level rise due to climate change is caused by the melting of land ice and \_\_\_\_\_.
  - A. Water volume expansion.
  - B. Water condensation.



C. Neither A nor B.

D. Both A and B.

4. The reason for my answer is:

- a. Water in the oceans will undergo this process due to the increase in average global temperatures.
- b. There will be an increase in water volume in the ocean as there will be consistently more rainfall near coastal areas.
- c. There will be more water in land than in the oceans as evident by the increased frequency of floods.
- d. Others: \_\_\_\_\_

## References

- Bloom, H. S. (2003). Using “short” interrupted time-series analysis to measure the impacts of whole-school reforms with applications to a study of accelerated schools. *Evaluation Review*, 27(1), 3–49.
- Boon, H. J. (2010). Climate change? Who knows?. A comparison of secondary students and pre-service teachers. *Australian Journal of Teacher Education*, 35(1), 104–120. Borckardt, J. J. Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O’Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. *American Psychologist*, 63(2), 77–95.
- Chang, C. H. (2014). *Climate change education: Knowing, doing and being*. Abingdon: Routledge.
- Chang, C. H., & Pascua, L. (2014). Uncovering the nexus between scientific discourse and school geography in Singapore students’ understanding of climate change. *Research in Geographic Education*, 16(1), 41–56.
- Chang, C. H., & Pascua, L. (in press). Singapore students’ misconceptions of climate change. *International Research in Geographical and Environmental Education* (in press).
- Cheam, J. (2012 (February 10)). More S’poreans care about climate change. *The Straits Times* Retrieved from <https://www.nccs.gov.sg/news/straits-times-more-s-poreans-care-about-climate-change/>.
- Check, J., & Schutt, R. K. (2012). *Research methods in education*. California: Sage Publications.
- Chi, M. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 61–82). New York, NY: Routledge.
- Chi, M., & Roscoe, R. (2002). The processes and challenges of conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 3–27). The Netherlands: Kluwer Academic Publishers.
- Cordero, E., Todd, A., & Abellera, A. (2008). Climate change education and the ecological footprint. *Bulletin of the American Meteorological Society*, 89(6), 865–872.
- Duit, R., & Treagust, D. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671–688.
- Duit, R. (1999). Conceptual change approaches in science education. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 263–282). Oxford: Elsevier Science Ltd.
- Dupigny-Giroux, L. A. L. (2010). Exploring the challenges of climate science literacy: Lessons from students, teachers and lifelong learners. *Geography Compass*, 4(9), 1203–1217.
- Farnsworth, C. H., & Mayer, V. J. (1984). An assessment of the validity and discrimination of the intensive time-series design by monitoring learning differences between students with different cognitive tendencies. *Journal of Research in Science Teaching*, 21(4), 345–355.
- Flowers, A. B. (2010). Blazing an evaluation pathway: Lessons learned from applying utilization-focused evaluation to a conservation education program. *Evaluation and Program Planning*, 33(2), 165–171.
- Gautier, C., Deutsch, K., & Rebich, S. (2006). Misconceptions about the greenhouse effect. *Journal of Geoscience Education*, 54(3), 386–399.
- Gottman, J., McFall, R., & Barnett, J. (1969). Design and analysis of research using time series. *Psychological Bulletin*, 72(4), 299–306.
- Hansen, P. (2010). Knowledge about the greenhouse effect and the effects of the ozone layer among Norwegian pupils finishing compulsory education in 1989, 1993, and 2005: What now? *International Journal of Science Education*, 32(3), 397–419.
- Harrington, J. (2008). Misconceptions: Barriers to improved climate literacy. *Physical Geography*, 29(6), 575–584.
- Henry, G. T., & Rubenstein, R. (2002). Paying for grades: Impact of merit-based financial aid on educational quality. *Journal of Policy Analysis and Management*, 21(1), 93–109. Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165–179.
- Lambert, J., Lindgren, J., & Bleicher, R. (2012). Assessing elementary science methods students’ understanding about global climate change. *International Journal of Science Education*, 34(8), 815–843.
- Lee, O., Lester, B., Ma, L., Lambert, J., & Jean-Baptiste, M. (2007). Conceptions of the greenhouse effect and global warming among elementary students from diverse languages and cultures. *Journal of Geoscience Education*, 55(2), 117–125.
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, 11(4), 357–380.
- Lin, H., & Lawrenz, F. (1999). Using time-series design in the assessment of teaching effectiveness. *Science Education*, 83(4), 409–422.
- Marston, D. (1988). The effectiveness of special education a time series analysis of reading performance in regular and special education settings. *The Journal of Special Education*, 21(4), 13–26.
- May, H., & Supovitz, J. A. (2006). Capturing the cumulative effects of school reform: An 11-year study of the impacts of America’s choice on student achievement. *Educational Evaluation and Policy Analysis*, 28(3), 231–257.
- Mayer, V. J., & Kozlow, M. J. (1980). An evaluation of a time-series single-subject design used in an intensive study of concept understanding. *Journal of Research in Science Teaching*, 17(5), 455–461.
- Mayer, V. J., & Lewis, D. K. (1979). An evaluation of the use of a time-series single-subject design. *Journal of Research in Science Teaching*, 16(2), 137–144.
- Mayer, V. J., & Rojas, C. A. (1982). The effect of frequency of testing upon the measurement of achievement in an intensive time-series design. *Journal of Research in Science Teaching*, 19(7), 543–551.
- McCaffrey, M. S., & Buhr, S. M. (2008). Clarifying climate confusion: Addressing systemic holes, cognitive gaps, and misconceptions through climate literacy. *Physical Geography*, 29(6), 512–528.
- Ministry of Environment and Water Resources (2006). *The Singapore green plan 2012* (2006 ed.). Singapore: Author.
- Monk, J. S. (1984). An examination of methods used to generate daily group scores from single-item-per-subject data collected in intensive time-series designs. *Journal of Research in Science Teaching*, 21(3), 315–324.
- Moon, S., Stanley, R. E., & Shin, J. (2006). Measuring the impact of lotteries on state per pupil expenditures for education: Assessing the national evidence. *Review of Policy Research*, 22(2), 205–220.
- Mower, T. (2012). Climate change in the curriculum: Are all young people informed and inspired? *Earth & Environment*, 8(1), 1–37.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Boston, MA: Pearson.
- Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25(1), 26–45.
- Österlind, K. (2005). Concept formation in environmental education: 14-year olds’ work on the intensified greenhouse effect and the depletion of the ozone layer. *International Journal of Science Education*, 27(8), 891–908.
- Papadimitriou, V. (2004). Prospective primary teachers’ understanding of climate change, greenhouse effect, and ozone layer depletion. *Journal of Science Education and Technology*, 13(2), 299–307.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). California: Sage.
- Punter, P., Ochando-Pardo, M., & Garcia, J. (2011). Spanish secondary school students’ notions on the causes and consequences of climate change. *International Journal of Science Education*, 33(3), 447–464.
- Rebich, S., & Gautier, C. (2005). Concept mapping to reveal alternative conceptions and conceptual change in a mock summit course on global climate change. *Journal of Geoscience Education*, 53(4), 355–365.
- Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology*, 31(1), 64–96.
- Shepardson, D., Niyogi, D., Choi, S., & Charusombat, U. (2009). Seventh grade students’ conceptions of global warming and climate change. *Environmental Education Research*, 15(5), 549–570.
- Tan, G. C. I. (2013). Voices in Singapore: Young people visioning their futures. In M. Robertson & S. Tani (Eds.), *Young people: Cross-cultural views and futures* (pp. 107–124). Victoria, Australia: Acer Press.
- Treagust, D. F. (2012). Diagnostic assessment in science as a means to improving teaching, learning and retention. *Proceedings of The Australian Conference on Science and Mathematics Education (formerly UniServe Science Conference)*. <http://openjournals.library.usyd.edu.au/index.php/ISSME/article/viewFile/6375/7014>
- Tyler, R. W. (1949, 2013). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.
- William, D., & Thompson, M. (2009). Integrating assessment with learning: What will it take to make it work? In C. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). New York, NY: Taylor & Francis Group.