2023-05-01

Incorporating Community Science In Improving Environmental Data Quality Through Model Based Reasoning Techniques

John Gilbert Olgin
University of Texas at El Paso

INCORPORATING COMMUNITY SCIENCE IN IMPROVING ENVIRONMENTAL DATA QUALITY THROUGH MODEL BASED REASONING TECHNIQUES

JOHN GILBERT OLGIN

Doctoral Program in Geological Sciences

APPROVED:

Deana Pennington, Ph.D., Chair

James Kubicki, Ph.D., Department Chair

Tom Gill, Ph.D.

Adriana Perez, Ph.D.

Marile Colon Robles, M.S.

Laura Serpa, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

For mom, dad and tita

INCORPORATING COMMUNITY SCIENCE IN IMPROVING ENVIRONMENTAL DATA QUALITY THROUGH MODEL BASED REASONING TECHNIQUES

by

JOHN GILBERT OLGIN, M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Earth, Environmental and Resource Sciences
THE UNIVERSITY OF TEXAS AT EL PASO
May 2023

ACKNOWLEDGEMENTS

Special thanks to my colleagues and collaborators at El Paso Community College, The University of Texas at El Paso, and NASA’s GLOBE program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Citizen science has been gaining in popularity since the later part of the 20th century, where laymen can contribute to the advancement of science by collaborating with researchers in various fields in science (Ifenthaler & Seel, 2013). Most are familiar with amateur astronomy (Edgett & Christensen, 1996; Graham et al., 2011) and ornithology (Benedetti et al., 2018) which have been popular for decades and have yielded important contributions in research. The explosion in Citizen science in recent times has contributed vast amounts of data that otherwise would not have been possible through traditional means (Auerbach et al., 2019; Fritz et al., 2019; Henderson et al., 2006). This is an overall success for both types of scientists: formal scientists who gain access to increasing amounts of data for their research, and Citizen scientists who benefit from learning how science works and relates to their daily lives. Making science more relatable for Citizen scientists furthers their curiosity of science and promotes the application of science within the Citizen (Ault et al., 2006; Jollymore et al., 2017).

Another good example is the Global Learning and Observations to Benefit the Environment (GLOBE) Program run by the National Center for Atmospheric Research (NCAR), University Corporation of Atmospheric Research (UCAR), the National Science Foundation (NSF) and the National Aeronautics and Space Agency (NASA). Its vision and mission is to build a global team to improve Earth's environment and promote academic advancement in environmental stewardship, literacy and discovery respectively. Beginning in the mid 1990's, GLOBE engaged with K-12 students conducting laboratory activities, known as protocols, in four principal themes: atmosphere, pedosphere, hydrosphere and biosphere. Results from these protocols were sent to GLOBE for scientists working in those areas to utilize the data for their respective investigations. The program grew in scope, and now data collection is open to a global audience of all ages using the GLOBE Observer app for smart devices. Data collected in a wider range of topics, such as cloud formation and mosquito habitats, offers scientists a greater volume and variety of data that can be utilized.

However, Citizen science data collection has had some drawbacks. Most common of these is lower data quality, where errors are introduced by the Citizen science participant. Examples include miscategorizing fauna which leads to an inaccurate count of species abundance in a particular region of study (Mitchell et al., 2017), or miscategorizing cloud formations due to the lack of formal training and practical field experience (Amos, Helen, personal communication, April 23, 2020). Problems lie with the lack of knowledge and expertise of the Citizen scientist, affecting their productivity in data collection. This has led scientists to question the reliability of these data and proceed cautiously when incorporating these data in their research (Mitchell et al 2017; Fritz et al, 2017; Vann-Sander et al, 2016; Jollymore et al, 2017; Amos et al, 2020). Data quality has been improved by designing Citizen science projects such that statistical analysis can be conducted to identify reliable data (Stylinski et al., 2020). During the 2017 solar eclipse it was necessary to collect temperature variation data from more than 20,000 GLOBE Observer participants yielding statistically significant results when corroborated with Mesonet data (Dodson et al., 2019). Cloud data taken between 2016 and 2019 show only 68% of cloud observations recorded cloud classifications (Amos et al., 2020). Further analysis of specifically dust emission data between January and April 2020 show a 74% of uncorroborated detection with satellite data of dust emissions globally (Amos et al., 2020; Dodson et al., 2019). Furthermore, the report from Amos et al. (2020) describes a comprehensive review of overall data collected across NASA's GLOBE Observer protocols (i.e. clouds, mosquito habitat, land cover and trees), indicating other data quality issues such as logistical errors. Knowledge and experience deficiencies need to be addressed to increase data integrity and enable Citizen scientists to engage effectively in science projects in their Citizen.

Most Citizen science projects provide some kind of online training for participants (Aye et al., 2019; MacDonald et al., 2018; Stylinski et al., 2020). For example, Archer et al. (2018) provided basic training in auditory analysis for Citizen scientists to study ultra-low frequency (ULF) waves when studying geomagnetic storms, and Aye et al. (2019) offered online training for Citizen scientists to practice identifying sites on the Martian surface for CO_2 fan deposits using data from HiRISE. The results show that the fan spread had a standard deviation of $5^{o} \pm 3^{o}$, which provided accurate enough results for other planetary scientists to use (Aye et al, 2019).

Interventions to improving data quality have been introduced, such as problem based learning (PBL), though with limited success (Mitchell, 2017). However, one approach appears may prove more effective. A learning theory known as model-based reasoning (MBR) has been developed and applied over the past two decades to facilitate complex reasoning tasks (Ifenthaler & Seel, 2013; Nersessian, 2009) . MBR is based on the notion that creation of external representations (diagrams, charts, or other visuals that provide a simplified model of the problem) invokes revision of internal mental models through time, leading to more accurate mental models. MBR has been demonstrated to be a key mechanism by which scientists reason about complex information (Ifenthaler & Seel, 2013; Jones et al., 2011). The hypothesis of this project is that training Citizen scientists using MBR will improve their understanding of physical processes which in turn will improve data quality. This hypothesis will be tested in an existing Citizen college service learning program that requires students to conduct Citizen science.

This dissertation will address the questions: *In what ways can MBR theory be applied to design training that will improve the capacity of Citizen scientists to collect and analyze high quality scientific data? To what degree does this training impact their ability to formulate complex understandings of environmental problems in their Citizen?* The project will be comprised of three research objectives focused in three sections, carried out with NASA's GLOBE Observer Citizen Science Program:

1 *Section 1*: An analysis of existing GLOBE data will be conducted, identifying where errors in data collection originate. At least three principal errors will be selected from this analysis;

2 *Section 2*: An online training regimen for participants will be developed that targets the selected errors; one based in traditional training techniques, the other rooted in MBR theory. The training will be implemented in an existing El Paso Citizen College (EPCC) service-learning course. Participants will be evaluated prior and after training to measure changes in knowledge, skills, and data quality compared with traditional training mechanisms;

3 *Section 3*: The training will then be evaluated and determine additional action to further develop the application of MBR to improve data quality.

The expected outcome is for those participants trained to record cloud and dust event data using MBR to show increased data collection accuracy and more complex formulations of scientific problem-solving experience in environmental processes. This will improve our understanding of how to more effectively train Citizen scientists. This knowledge can be applied in other Citizen science contexts.

ERROR ANALYSIS

INTRODUCTION

Citizen science has been gaining in popularity since the later part of the 20th century, where laymen can contribute to the advancement of science by collaborating with researchers in various fields in science (Barrutia et al., 2021; Callaghan et al., 2021; de Sherbinin et al., 2021; Ifenthaler & Seel, 2013; Pernat et al., 2021). Most are familiar with amateur astronomy (Edgett & Christensen, 1996; Graham et al., 2011) and ornithology (Benedetti et al., 2018) which have been popular for decades and have yielded important contributions in research. The explosion in citizen science in recent times has contributed vast amounts of data that otherwise would not have been possible through traditional data collection methods (Auerbach et al., 2019; Avard & Clark, 2001; Callaghan et al., 2019; Follett et al., 2019; Henderson et al., 2006; Pernat et al., 2021) . This is a win-win for both types of scientists: formal scientists who gain access to increasing amounts of data for their research, and citizen scientists who benefit from learning how science works and relates to their daily lives. Making science more relatable for citizen scientists furthers their curiosity of science and promotes the application of science within the community (Ault et al., 2006; Barrutia et al., 2021; Callaghan et al., 2019; de Sherbinin et al., 2021; Jollymore et al., 2017; Koffler et al., 2021). These types of investigations offer easily attainable objectives, such as classification of physical characteristics, that allow citizen scientists to thrive in these collaborations.

Specific large-scale successful citizen science programs in the USA that have helped scientists with their research include Ebirds, which has brought synergy between citizen science participants and researchers in ornithology, yielding data in avian biodiversity (Sullivan et al., 2014); and the Community Collaborative Rain, Hail and Snow Network (CoCoRAHS), which has citizen science participants recording different types of precipitation data for the research and operational community (Reges et al., 2016). One of the nation's largest citizen science networks, CoCoRAHS has collected over 31 million daily reports from over 37,000 participants in a span of 17 years (Reges et al., 2016). Impacts from these types of programs have greatly improved science and science awareness of the general public (Tierney et al., 2020).

Another example is the Global Learning and Observations to Benefit the Environment (GLOBE) Program run by the National Center for Atmospheric Research (NCAR), University Corporation of Atmospheric Research (UCAR), the National Science Foundation (NSF) and the National Aeronautics and Space Agency (NASA). Its vision and mission are to build a global team to improve Earth's environment and promote academic advancement in environmental stewardship, literacy and discovery (Amos et al., 2020; Enterkine et al., n.d.; Kohl et al., 2020; Robles et al., 2020; Smolleck et al., 2006). Beginning in the mid 1990's, GLOBE engaged with K-12 students conducting laboratory activities, known as protocols, in four principal themes: atmosphere, pedosphere, hydrosphere and biosphere. Results from these protocols were sent to GLOBE for scientists working in those areas to utilize the data for their respective investigations. The program grew in scope, and now data collection is open to a global audience of all ages using the GLOBE Observer app for smart devices. Data collected in a wider range of topics, such as cloud formation and mosquito habitats, offers scientists a greater volume and variety of data that can be utilized.

However, citizen science data collection has had some drawbacks. Most common of these is lower data quality, where errors are introduced by the citizen science participant. For most of these investigations, participants were often prompted to classify and identify certain physical characteristics, such as types of fauna (Mitchell et al., 2017). Issues arose when participants attempted to accurately complete these tasks. Problems lie with the lack of knowledge and skill of the citizen scientist, affecting their productivity in data collection. For example, errors in classifying dust storms are an issue with citizen scientists as well as some professional weather professionals (Ardon-Dryer et al., 2022). Mitchell et al. (2017) reported that misclassified fauna led to an inaccurate count of species abundance in a particular region of study. Similarly, scientists attempting to use data collected by citizen scientists using the GLOBE Observer protocol Mosquito Habitat have encountered errors in identification or classification of mosquito larvae (Low et al., 2021). They modified data acquisition techniques to minimize classification errors. One of the solutions was to eliminate the actual larvae classification and focus on data that was less prone to user error, improving data quality and reliability overall (Cuzzolino et al., 2019; Low et al., 2021; Lukyanenko et al., 2019). Data quality has also been improved by designing citizen science projects such that statistical analysis can be conducted to identify reliable data (Stylinski et al., 2020). This works with certain types of experiments, but not all. During the 2017 solar eclipse it was necessary to collect temperature variation data from more than 20,000 GLOBE Observer participants in order to yield statistically significant results (Dodson et al., 2019). Analysis of GLOBE dust emission data between January and April 2020 show 74% of uncorroborated detection of dust emissions globally with satellite data (Amos et al., 2020; Dodson et al., 2019). GLOBE cloud data taken between 2016 and 2019 show only 68% of cloud observations recorded cloud classifications (Amos et al., 2020). The GLOBE Observer Clouds protocol faces similar identification and classification issues (Amos, Helen, personal communication, April 23, 2020; Colon-Robles et al., 2019). Incomplete data sets, misclassification of clouds and misidentification of obscured skies as overcast ones were some of the issues with data submitted to GLOBE. Furthermore, the report from Amos et al. (2020) describes a comprehensive review of overall data collected across NASA's GLOBE Observer protocols (i.e. clouds, mosquito habitat, land cover and trees), indicating other data quality issues such as logistical errors. This has led scientists to question the reliability of these data and proceed with caution when incorporating these data in their research (Aceves-Bueno et al., 2017; Amos et al., 2020; Fritz et al., 2017; Jollymore et al., 2017; Lukyanenko et al., 2019; Vann-Sander et al., 2016).

These examples point to a large problem with citizen science data quality. It is critical to obtain more accurate data to increase reliability and confidence in using the data, which depends on more effective training of citizen scientists. Knowledge and experience deficiencies need to be addressed to increase data integrity and enable citizen scientists to engage effectively in science projects in their community, while also maintaining interest in conducting investigations without cumbersome effort in learning a new skill set. Data collection training needs to be embedded within sufficient theoretical training on the phenomena of interest to enable critical thinking during data collection (Callaghan et al., 2021; de Sherbinin et al., 2021) and foster further curiosity in collaborating with scientists. Yet without a better understanding of why these errors are made it is not possible to design theoretical training that targets the source of the errors. There has not been a comprehensive analysis of sources of errors to enable effective training design. This article reports on a comprehensive investigation of errors in the GLOBE Observer Cloud dataset as a precursor to developing and testing new training approaches.

METHODOLOGY

Datasets selected in this study were taken from the NASA GLOBE observer website for the year 2020. Five hundred datasets were randomly selected and downloaded through a script. Datasets included information about the user, site (latitude, longitude, and elevation), and date along with observations made by the user following the GLOBE Observer Cloud Protocol: 1) sky obscurations (fog, heavy rain or snow, sand, spray, smoke, dust, haze or volcanic ash); 2) categorization of any clouds into one of four altitudinal categories and classification of the cloud type within each altitudinal category — very high-level airplane contrails, high-level (cirrus, cirrocumulus, or cirrostratus), mid-level (altostratus or altocumulus), or low-level (fog, nimbostratus, cumulonimbus, stratus, cumulus, or stratocumulus); 3) atmospheric conditions such as an estimate of percent cloud cover and visual opacity (opaque, translucent, or transparent); and 4) surface conditions, including identification of snow/ice, standing water, muddy, dry ground, leaves on trees, and raining/snowing. Multiple cloud types at the same or different levels can be indicated by the participant. Participants take up to six photos (i.e. north, south, east, west, zenith and nadir photos). Only datasets with a complete set of photos were selected for this investigation so that the entire field of view of the user was captured.

Each set of six photos was analyzed by the investigator, categorized according to the four altitudinal categories established by GLOBE Observer and the observed cloud types classified. Each observation was analyzed to compare user classifications with those of the investigator. Grounded theory was used to develop a list of errors and possible explanations of their source (Table 1) (Charmaz & Thornberg, 2021; Chun Tie et al., 2019; Levitt, 2021). Grounded theory is a systematic method for generating a coding scheme from the data as it is being analyzed rather than developing a coding scheme in advance. The first column of Table 1 shows the classification correctness type. Users either completely misclassified cloud types (code 0), completely correctly classified cloud types (code 1), partially correctly classified cloud types in multicloud assemblages (code 2), or the correct cloud types were unable to be verified by the investigator (unknown - code 3). Those with code 2.1 identified fewer cloud types than were present in the data (under counted), although those identified were correct. Those with code 2.2 identified more cloud types than were present in the data (over counted), although some of the clouds were correctly classified. Code 3 represented insufficient data in the photos to corroborate the cloud classification. This code was primarily used if there were physical obstructions blocking the view, sky obscurations made classification from photos difficult, or photos were blurry. The second column in Table 1 represents contextual nuances in the data that may have impacted misclassifications. A Contextual code of 1 indicated any type of misclassification error (Correctness Types 0, 2, or 3). Mixed cloud types with similar characteristics, such as cumulus-type clouds from the mid-level combined with low-level clouds, would be a key indicator of "Dismissed data" (code 2). Over generalized classification (code 3), where users possibly "second guessed" their judgment on cloud types, is indicated when a cloud type is generalized (i.e., cumulus cloud instead of stratocumulus). The third column in Table 1 (Comments) represents a free text description of observed classification errors, adding additional nuance beyond Contextual Type codes. For example, high-level clouds being mistaken for low-level clouds would indicate users have limited understanding of the cloud conditions to make a correct assessment. Improper data collections, such as blurry photos, would be labeled in this third column.

Correctness and Contextual Type codes were then compiled across all observations and results plotted. Comments were further analyzed using an iterative inductive approach to identify common patterns in errors.

Table 1: Rubric used to analyze and characterize data errors. Grounded theory was used to develop the rubric.

Correctness Types	Contextual Types	Comments
0 Incorrect classification 1 Correct classification 2 Partially correct classification 2.1 Undercounted 2.2 Overcounted 3 Indeterminate	1. Misclassification of cloud types 2. Dismissed data: Possible lack of user confidence in classification 3. Over generalized classification	General commentary on classification process, notable conditions of the data, spot any difficulties in classification of cloud types.

RESULTS

Of the 500 samples analyzed, 150 (34%) were correctly classified, 180 (36%) were misclassified, 125 (25%) were partially correctly classified, and 45 (9%) were indeterminate (Figure 1). Those partially correctly classified included 65 (13%) that were undercounted and 60 (12%) that were overcounted.

Samples were subdivided and analyzed for each of the four altitudinal categories (contrails, high-, mid-, and low-level clouds). Results show varying degrees of misclassification and unknown classifications across these groups (Figure 2), with contrails showing the lowest number of correct classifications (20%), low-level clouds the most correctly classified (60%, and high- and mid-level clouds having similar indeterminate levels of correct classification (38% and 37%, respectively). Results indicate that accuracy increases as the altitude of clouds decreases. Beyond this observation, few obvious patterns are present in the data. Contrails had, by far, the highest number of partially correct samples that were overcounted (36%, compared with at or less than 10% at the other three altitudes. In contrast, contrails had the lowest number of partially correct samples that were undercounted (less than 5%). Contrails also had the highest number of samples whose cloud type was indeterminate (~30%).

Contextual analysis (Table 1) of the samples shows an increase in over generalized classifications with increase in cloud altitude, from 28% from low level clouds to 47% with contrails. Dismissed data also increased with increasing cloud altitude, from 21% in low level clouds to 38% in contrails.

The analysis of the comments about each data set identified keywords indicating possible causes for classification inaccuracies; the most common five keywords are shown in Table 2.

Table 2: Top five most common comment keywords.

Keyword	Description
Mix	Overlap of various cloud types, possibly causing confusion and difficulty of determining cloud types properly.
Difficult	Issues with data acquisition, such as obstacles interfering with capturing clouds, overcast conditions proving an issue with confirming details of clouds to determine cloud type.
Missing	Incomplete accounting of all cloud types while other cloud types were correctly accounted for.
Conditional	Confusion of cloud types; mistaking one cloud type for another (i.e. altocumulus for cumulus clouds).
Data acquisition error	Photos that were out of focus, blurry

An overall assessment of the results from data analysis indicates:

- Two-thirds of the data are categorized as incorrect, partially incorrect or indeterminate; only one-third were correctly categorized by participants.
- Accuracy in cloud classification increases as cloud altitude decreases.
- Clouds at higher altitudes prove difficult for participants to identify correctly, especially when it comes to more detailed characteristics that help in differentiating contrails and the cirrus family of clouds.
- Misclassification of cumulus type of clouds across all altitudinal categories is noted.

From this study, we identify three major errors impacting data quality: 1) “Contrail Confusion”; 2) “Cumulus Conundrum” and 3) “Coexisting Clouds”.

Contrail Confusion: Cirrus cloud family vs contrails. The nature of cirrus (high level) clouds, such as their “whispy” nature and dispersed appearance, prove it to be somewhat challenging to correctly classify these, as some contrails have similar features. Contrail features will also disperse over time and can blend in with the background cirrus clouds should they be present. Another issue with contrails is that the appearance of them may come after the participant has submitted their initial observations, possibly further confusing the observer on including these data in their submission to GLOBE Observer.

Cumulus Conundrum: Cirrocumulus vs. Altocumulus vs. Cumulus clouds. Here the issue appears to lie with depth perception. The aforementioned cloud types have the classic “cloud bulging" or "fluffy" characteristics; however, some high-, mid- and low-level clouds have these same characteristics. Oftentimes high- and mid-level clouds are confused with the low-level cumulus clouds, as it appears observers cannot make the distinction between cloud levels. A better understanding of depth perception could be a possible focal point for better classifying these cloud types properly.

Coexisting Clouds: Overlaying cloud types. It is rare for solely one type of cloud formation to exist at any given time. Especially during storms, a variety of cloud types will emerge as a storm builds within the viewing region. Coexisting cloud types are easily overlooked or misclassified due to the convergence of characteristics.

DISCUSSION

Previous studies of cloud classification efforts by citizen scientists have noted observer errors such as incomplete data sets and misclassification of cloud types (Amos et al., 2020; Dodson et al., 2019; Robles et al., 2020). However, specific errors, as identified in this article (i.e. "Contrail Confusion", "Cumulus Conundrum" and "Coexisting Clouds"), have not explicitly been described in the literature but have been observed by NASA GLOBE scientists in the field (Amos, Helen, personal communication, April 23, 2020). As the results of the study showed, low level clouds were classified more successfully than those at higher altitudes. A recent study by Dodson et al. (2023) also highlights the difficulty of participants classifying high level clouds. The partially correct classifications also decreased as cloud altitude decreased. Dodson et al. (2022) found the inverse – that cloud classification from satellite imagery using machine learning methods was more accurate for high level clouds and contrails. Combining the findings of these two studies possibly suggests that classifiers, whether human or machine, are better able to identify cloud structures that are closer to them, which is somewhat intuitive. It also suggests that the technologies could potentially be combined in some way such that ground based observations are emphasized in classifying lower clouds while satellite based observations are emphasized in classifying higher altitude clouds. Such integration of disparate data could potentially improve classification accuracy. However, Dryer et al. (2023) found that identification of dust storms from integrated data from diverse source remained problematic.

Incorrect classifications, with the exception of low level clouds, increased somewhat with lowering altitude, possibly hinting at some common cloud structures that participants were having difficulty with. Figure 2a and 2b may also indicate a temporal factor to difficulties in cloud classification, as high level clouds alter their formation on longer time scales, especially contrails. It is possible that these points could offer some insight on how to structure effective training that addresses these issues.

Based on the findings of this article, a key issue that needs to be addressed during training is the difficulty of spatial and temporal representation of complex weather phenomena. Spatial representation is key in identifying and studying various meteorological and geophysical processes – especially atmospheric and surface processes (Gold et al., 2018; Johnson & McNeal, 2022; McLaughlin & Bailey, 2022; Sezen-Barrie et al., 2022). Research in geophysical education by Newcombe and Shipley (2015) provided an organizational structure, or framework, to further analyze and classifying structural characteristics in geoscience. This relatively novel framework could be applied more readily to study solid and fluid features covered in the geosciences, such as atmospheric and surface processes. This framework would guide students to analyze static and dynamic characteristics in these features, further developing spatial and temporal skills. To date, there have been limited efforts beyond Newcombe and Shipley (2015) reported to address the issue of spatial thinking in atmospheric science education research (Annis & Nardi, 2021; McLaughlin & Bailey, 2022; P. McNeal et al., 2018; P. M. McNeal & Petcovic, 2020) addressed a deficiency of a similar framework in atmospheric science that could prove useful in reducing cloud misclassification. Knowledge of cloud types brought through training in tandem with disembedding techniques, being able to observe and recognize distinct features unique to each cloud type (P. McNeal et al., 2018; P. M. McNeal & Petcovic, 2020), can greatly improve data quality of cloud classification.

One approach to overcoming spatial representation issues is to improve students' conceptual understanding of weather processes to enable critical thinking about their observations. Cervato et al. (2018) identify addressing misconceptions, pre-conceptions, partially correct conceptions and naïve conceptions as a grand challenge in atmospheric (and other earth) sciences. Barruita et al. (2019) also addresses misconceptions in weather phenomena, such as rain fall in relation to the water cycle. One point in particular that Cervato et al. (2018) focuses on is the use of models for predictions and atmospheric study. Research in cognitive science has long demonstrated that scientists rely heavily on simplified models to reason about complex processes and that the process of modeling invokes conceptual change (Nersessian, 2009). However, modeling phenomena to better understand conceptually physical processes using numerical and analytical models is difficult for novices to navigating through. Cervato et al. (2018) highlights the need for more effective utilization of models as learning tools. Schwarz et al. (2009) demonstrated the use of student-made illustrative models that improved student engagement and promoted better comprehension of science topics. Employing such an approach toward improving students' conceptual understanding of weather processes could conceivably improve critical thinking skills and decrease uncertainty in understanding atmospheric phenomena, which in turn would increase cloud identification.

To improve the quality of data collected by citizen scientists, modified training methods have been recommended along with better communication between scientists and citizen scientists regarding project objectives (Balázs et al., 2021; McLaughlin & Bailey, 2022). Existing training measures such as those introduced in GLOBE observer app protocols are excellent yet apparently insufficient to support accurate cloud classification. This is an unavoidable issue as in a recent study related to cloud classification shows that low level clouds are easier to classify than those higher up (Dodson et al., 2022). Participants may not have the techniques down to classify cloud types effectively, but prove to have an interest in participating in cloud classification challenges that GLOBE hosts (Dodson et al, 2022; Dodson et al, 2019). However, there is promise in offering effective training for citizen scientists.

McLaughlin and Bailey (2022), Gold et al (2018), and Uttal et al (2013) have demonstrated that targeted training vastly improves student's spatial classification of geophysical features in general. This could be focused and extended to atmospheric features. Furthermore, Kuhn et al. (2022) addresses instructor engagement with citizen scientists to fortify existing training methods ensuring success in data collection. Following this approach, inclusion of satellite data that is available through GLOBE could further enhance training overall. GLOBE Observer participants are emailed satellite data that GLOBE was able to match during their observations. Incorporating this into a training regimen offers not only further understanding on learning about clouds, but fortifies engagement and increases participant interest in continuing with data collection. This type of intervention can complement targeted training in areas where data quality needs further attention for improvement.

These recommendations may put an extra burden on citizen scientists; a learning curve that may dissuade interest. However, an effective training design will not only help participants to collect better data but increase interest and curiosity as their data becomes more utilized in scientific publications.

A vast expertise is not needed for GLOBE citizen scientists classifying clouds; however, a sufficient level of understanding is not an extreme demand put on those participating. A recent study on dust storm classification pointed out that even experts found difficulty in classifying them, confusing dust storms with haze or pollution (Dryer et al, 2022). It is only reasonable that the guidance for citizen scientists toward better data quality be effective, interactive, and insightful in order to further increase their interest in participation. Therefore, the need for modified and effective training for citizen scientists, especially in cloud data collection, is paramount to establish better rapport with science collaborators and offer more robust and accurate data for future use.

LIMITATIONS OF WORK

This investigation was limited by basic issues that impacted validation of participant classifications including:

- Participant's improper execution of the observation protocols;
- Confirmation of contrails was questionable, as the appearance of contrails can occur moments after initial observations are taken, as well as disappear out of sight of the observer and therefore cannot be accounted for in the data;
- Obstacles, such as buildings and trees, hamper proper identification and validation of data sets by the investigator; and
- Blurry photos also increase difficulty in validating data.

The study is also limited by the use of a single investigator to “correctly” classify images, which probably introduced error into the results. Use of multiple investigator classifications and comparison across these would reduce this error.

The findings of this study are limited to the case study: cloud classification by citizen scientists using the GLOBE app. More case studies in different contexts and cross case analysis would be needed to generalize any of these findings. Such generalization is needed to address the issue of data quality in citizen science projects in a comprehensive way.

FUTURE WORK

Modified and additional online training modules have been developed using the approach of student-made illustrative models, specifically concept maps Ifenthaler, D., & Seel, N. M,. 2013), that target the three identified misclassification errors. The modules are currently being tested in undergraduate atmospheric science classes.

CONCLUSION

This article investigated the source of classification errors in GLOBE cloud data collected by citizen scientists. Some of the takeaways from this study can be summarized by the following:

- High-, mid- and low-level clouds all share cumulus and stratus characteristics. The difficulty in deciphering high-, mid- or low-level clouds with those characteristics appears to be linked to the altitude of clouds;
- Low-level cumulus clouds are easier for participants to classify;
- High altitude contrails are commonly misclassified;
- Cloud misclassification often occurs when there are multiple types overlapping in the sky;
- Co-existing cloud types coupled with cloud elevation determination issues are types of spatial representation issues that have previously been identified in the solid earth science education literature but have been minimally investigated in the fluid earth sciences such as atmospheric science; and

- Resolving these issues will depend on improving citizen scientist's understanding of weather processes to enable critical thinking about their observations.

Key strategies for data quality improvement can be developed based on these conclusions. As these errors depend on the user to formulate a basic understanding of cloud types and weather processes that produce them, a training approach could be introduced to better develop citizen science understanding and increase data collection confidence. Future work to improve data quality will be to develop and test training methods that specifically target these issues.

ACKNOWLEDGMENTS.

Thanks to the collaboration and support from the NASA GLOBE Observer program for this study.

DATA AVAILABILITY STATEMENT.

All GLOBE data can be accessed at the following link: https://www.globe.gov/

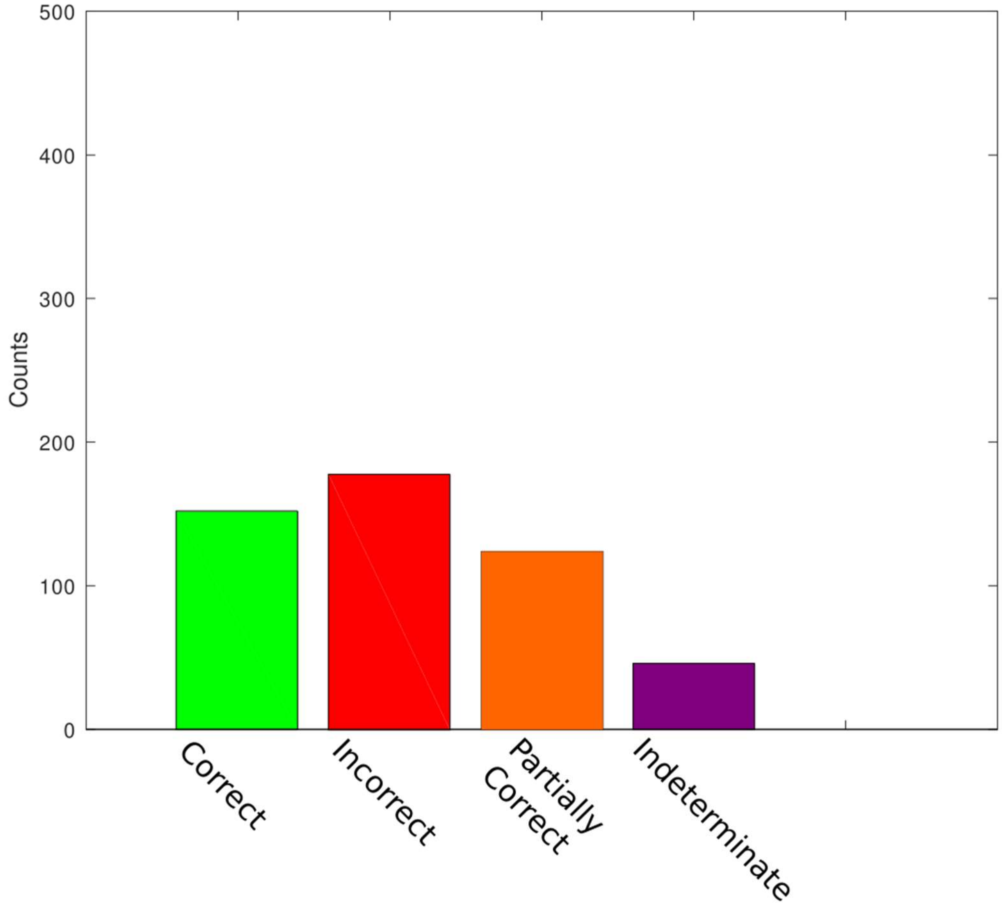


Figure 1.1: Results from 500 samples of cloud classifications: 150 samples correct (green), 180 incorrect (red), 125 partially correct (orange) and 45 indeterminate (purple).

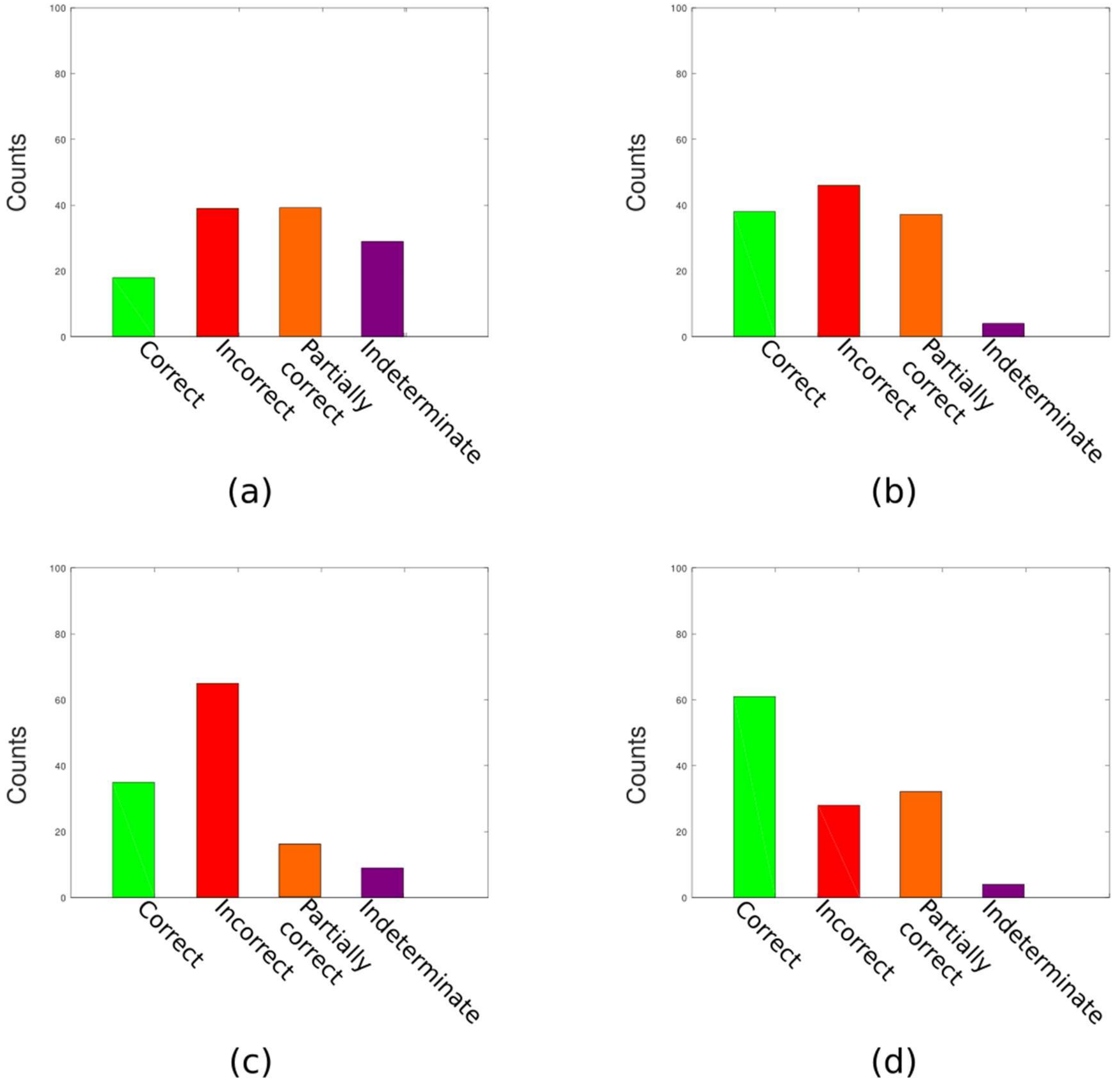


Figure 1.2: Classification rating results broken down for each cloud type category based on the four correctness types: contrails (a), high-level clouds (b), mid-level clouds (c), and low-level clouds (d).

MODEL BASED REASONING (MBR) TRAINING

INTRODUCTION

The latter part of the 20th century saw an explosion of citizen science contributions to the sciences, from amateur astronomy to ornithology (Ifenhaler & Seel, 2013; Graham et al, 2011). Data collection by citizen scientists benefits professional scientists by providing a volunteer workforce that can contribute more numerous observations, in more places and times, than could be collected by independent investigators. Examples of scientific investigations being enhanced by citizen science include the integration of water monitoring data by Canadian government agencies and citizen science groups to help improve the computer modeling of water resources in Canada (Deutsch et al, 2021); and citizen science monitoring of water quality and safety in Flint, MI, where they provided valuable data that further corroborated results gathered from government agencies (Peplow, 2018). Potential benefits to science are clear. There are also benefits for citizen scientists. Air quality projects that involved students in high school demonstrated the potential of collected data to be useful for scientists, but also serve as an effective learning tool that engages the curious minds (Lepenies and Zakari, 2021). During the COVID-19 lockdown, citizen science participation allowed secondary education students in Australia to continue with their virtual learning and provided scientists additional data in the process (Van Haeften et al, 2021). Post-pandemic efforts include NASA's GLOBE Air Quality Campaign that continues to benefit from the increased participation in data acquisition that began during the pandemic. The benefits of citizen science allow scientists to gain additional data and provide multiple learning experiences for participants to better familiarize with science at large.

Despite the valuable data contributions from citizen scientists, low data quality has become an issue. Errors are introduced by the citizen science participant in numerous ways, including spatial and temporal data biases (Low et al, 2021), poorly executed protocols and lack of adherence to instructions for data collection (Langenkamper et al, 2019; Hunter et al, 2012). This has led scientists to question the reliability of these data and proceed with caution when incorporating these data into their research (Low et al, 2021; Amos et al, 2020; Fritz et al, 2017; Jollymore et al, 2017; Mitchell et al 2017). Amos et al. (2020) describe a comprehensive review of data collected across NASA's Global Learning and Observations to Benefit the Environment (GLOBE) Observer protocols (i.e. clouds, mosquito habitat, land cover and trees), indicating data quality issues such as logistical errors. GLOBE cloud data taken between 2016 and 2019 show only 68% of cloud observations recorded cloud classifications (Amos et al., 2020). Further analysis of specifically dust emission data between January and April 2020 show a 74% of uncorroborated detection of dust emissions globally with satellite data (Amos et al., 2020; Dodson et al., 2019). Of these many potential sources of error, some are related to the lack of theoretical expertise of the citizen scientist about the phenomena being observed, affecting their capacity to critically assess their observations. An example is misclassification of cloud formations (Amos, Helen, personal communication, April 23, 2020). Many of these sources of error lend themselves to training interventions (Vohland et al, 2021). Data collection training can improve the data quality, allowing for sufficient theoretical thinking that leads to critical thinking with conducting observations (Callaghan et al, 2021; Sherbinin et al, 2021).

An analysis of data quality of GLOBE cloud observations was performed, and results revealed that out of 500 studied observations, 70% of those observations had at least one type of classification error (Olgin and Pennington 2022, unpublished data). Further analysis of incorrect classifications identified three common issues that could potentially be addressed through targeted training (Olgin and Pennington 2022, unpublished manuscript):

> 1) Contrail Confusion – participants confusing natural, high-level cloud formations with anthropogenic cloud formations (i.e. aircraft exhaust);
>
> 2) Cumulus Conundrum – issues with differentiating cloud type based on cumulus cloud characteristics among high, mid and low level cloud formations;
>
> 3) Coexisting Clouds – multiple cloud types prove confusing in properly classifying cloud formations.

This article reports on a citizen scientist training intervention that targeted classification errors made by citizen scientists using the GLOBE Observer application, applying a learning theory known as model-based reasoning (MBR; Ifenthaler & Seel, 2013; Nersessian, 2009). MBR is based on the notion that creation of external representations (diagrams, charts, or other visuals) that provide a simplified model of the problem invokes revision of internal mental models through time, leading to more accurate mental models. MBR has been demonstrated to be a key mechanism by which scientist's reason about complex information (Ifenthaler & Seel, 2013; Jones et al., 2011). This study introduced MBR to the training regime for a group of GLOBE Observer participants and tested it against those undergoing a traditional training regime.

METHODS

The three primary errors identified in the prior study of cloud data quality in GLOBE (Olgin and Pennington 2022, unpublished manuscript) were used to design MBR training modules to supplement the training provided by GLOBE. A heuristically attained approach to understanding physical processes, MBR provides a flexible and versatile framework to understand the dynamics of any process. This approach takes an inductive reasoning path that is more efficient and allows for productive cognitive growth. MBR methods include using a wide variety of external representations to invoke cognitive change; this study used concept maps (Ifenthaler & Seel, 2013; Jones & Reid, 2001; Liu & Stasko, 2010). Concept maps have been widely used in the science education Citizen to explore and assess students' conceptual understanding of key scientific concepts. Key words are introduced that are related to a particular topic and participants are prompted to make connections, or links, between those key words that express their understanding of process (Figure 2.1). The combination of two keywords and the directed link form a proposition (for example, Increasing Winds contributes to Dust Emissions). We hypothesized that incorporating concept mapping techniques into the cloud classification training process could provide a means to better develop their theoretical understanding of cloud formation processes, enabling better observations in the field.

For each of the three identified issues (Contrail Confusion, Cumulus Conundrum, Coexisting Clouds), seed questions were developed to target specific learning outcomes and prompt participant generation of relevant concept maps (Table 2.1). The instructor utilized a concept map rubric, to which participant concept maps could be compared. Two training curricula were developed: 1) a traditional training curricula leveraging the existing online GLOBE training materials (Figure 2.2); and 2) an MBR training curricula identical to the traditional curricula with the additional incorporation of concept mapping activities. The GLOBE training materials are offered through the GLOBE Observer app and online and include instruction for classifying cloud types, distribution and quantity. The traditional and MBR curricula are summarized in Table 2.1.

Table 2.1 List of three type error classifications gathered from observational data.

Error types	1. Contrail Confusion	2. Cumulus Conundrum	3. Coexisting Clouds
Seed prompts for MBR concept maps	Q1. Identify the different progression of contrail formation. Q2. What conditions influence contrail formation? Q3. What key features separate natural versus anthropogenic clouds? Q4. How can contrails resemble natural cirrus clouds?	Q5: What are the key characteristics that differentiate altocumulus, cirrocumulus, and cumulus/altocumulus clouds? Q6: What atmospheric and ground conditions are unique to these forms of cumulus cloud formations?	Q7: Identify key conditions that are different between obscurations and overcast skies. Q8: What are at least some conditions that might overlap between obscuration and overcast conditions?

Table 2.2 A comparative summary of traditional versus MBR training

Traditional training	Model Based Reasoning (MBR)
1 Participants perform online training through modules for identifying clouds types	4 Participants perform online training through modules for identifying clouds types
2 Assessments are conducted for reflection and auto-feedback from training modules	5 Assessments are conducted for reflection and auto-feedback from training modules
	6 Participants create concept maps that address questions in table 2.1.
3 Participants conduct outdoor observations in the field and submit results to NASA's GLOBE database through app	7 Participants conduct outdoor observations in the field and submit results to NASA's GLOBE database through app

Two cohorts of study participants, both in the summer of 2022, were established to test the efficacy of traditional training accompanied with MBR training when compared to traditional training alone. Both cohorts were freshman/sophomore level students from the El Paso Citizen College (EPCC) registered for an introductory atmospheric science course that included a module conducting citizen science with the GLOBE Observer app. Human subjects research oversight and approval was obtained from the EPCC Institutional Review Board (1699560-1). Participation in the research was voluntary, consensual, and anonymous.

Each cohort was separated into two groups – those receiving traditional training alone and those receiving additional MBR training (Table 2.3). The eight in the MBR group represent those who completed all MBR training modules.

Table 2.3 Summer 2022 Training Cohorts

Group type	Number of active participants	Number of observations
Traditional	11	80
MBR	7	56

Each group conducted observations in the field using the Clouds Protocol on the GLOBE Observer app. The Clouds observation protocol specifies the following observations: 1) sky obscurations (fog, heavy rain or snow, sand, spray, smoke, dust, haze or volcanic ash); 2) categorization of any clouds into one of four altitudinal categories and classification of the cloud type within each altitudinal category — very high-level airplane contrails, high-level (cirrus, cirrocumulus, or cirrostratus), mid-level (altostratus or altocumulus), or low-level (fog, nimbostratus, cumulonimbus, stratus, cumulus, or stratocumulus); 3) atmospheric conditions such as an estimate of percent cloud cover and visual opacity (opaque, translucent, or transparent); and 4) surface conditions, including identification of snow/ice, standing water, muddy, dry ground, leaves on trees, and raining/snowing. Multiple cloud types at the same or different levels can be indicated by the participant. The app includes functionality for taking six photos at the time of each observation (north, south, east, west, zenith and nadir photos). All observations were then submitted to the GLOBE cloud database by the participant. Each location where the observation was taken was cataloged with a site ID by GLOBE that participants recorded and submitted to the investigator.

Data associated with each site ID was downloaded from GLOBE by the investigator for analysis. In addition to the observations above, downloaded data included information about the user, the site (latitude, longitude, and elevation), and the date observations were made.
Photos were used to assess cloud classification accuracy for each set of data. A count of the number of correct cloud classifications was recorded for each cloud type, for each cohort and group. Cohorts were combined and accuracy compared between the traditional and MBR groups using basic statistics, boxplots and histograms.

The MBR group's concept maps were assessed for completeness and accuracy, using expert maps generated by the investigator as a reference. A rubric developed by the National Park Service (https://www.nps.gov/grsm/learn/education/classrooms/upload/Concept-Map-Scoring-Rubric.pdf) was used to assess the concept maps. Each concept map was assessed according to three criteria: organization, content-concepts-terminology, and connections-relationships. A score was assigned for each criterium in the rubric; 0 being the lowest score and 4 the highest. The three scores were summed to generate a final score for each concept map developed by each MBR participant; the highest possible final score was 12. Those scores were compared with the corresponding observation accuracy using a scatterplot.

The percentage of correct classifications was calculated for the two groups. The MBR group was subdivided by grouping students with the highest (8-10), middle (4-7), and lowest (0-3) concept map scores. The percentage of correct classifications was calculated for the three subset groups. Boxplots were constructed for all five groups. Histograms were created comparing the accuracy of the two unsubdivided groups (traditional and MBR) and comparing the traditionally trained group with the three subdivided MBR groups (high, middle, and low scoring concept maps). All five groups were further subdivided by cloud type: high-, mid-, and low-level clouds or contrails. Histograms were constructed comparing classification accuracy by cloud type. A Mann-Whitney U test was used to quantify the statistical significance of differences across independent groups. This is a non-parametric test appropriate for non-normally distributed small samples.

RESULTS

A comparison between concept map scores and classification accuracy indicates there is no significant correlation between these data, with a calculated R^2 of 0.06 (Figure 2.3). Mean classification accuracy of traditional and MBR training groups were 77.9% and 78.3% respectively, with the MBR group scoring approximately 0.4 percentage points higher than the traditional group (Figure 2.4). The MBR high and low scorers performed better than the traditional group by approximately 0.9 and 1.0 percentage points respectively (Figure 2.5). Mid scorers from the MBR group performed on par with the traditional group and underperformed the low and top MBR groups (Figure 2.5). Combining the low and top scores and comparing with the traditional group shows an almost 1% increase in accuracy in cloud classifications. Results of the Mann Whitney U test indicate that in comparison with the traditional group, there was no significant difference in classification accuracy for any of the MBR groups (Table 2.4).

Table 2.4 Results from statistical analysis using the Mann Whitney U non-parametric test of rank sum values across independent samples.

Groups	Z value	P value
Traditional / MBR	0.08	0.94
Traditional / MBR Low	-0.08	0.94
Traditional / MBR Mid	-0.08	0.94
Traditional / MBR High	-0.13	0.90

The Mann-Whitney U test is a ranked sum test of differences in median values. Boxplots indicate the median accuracy value of the traditionally trained group was higher than the MBR full group by 2.7 percentage points (79.2% and 76.5%, respectively; Figure 2.6). Median values for the MBR subgroups were 73.8% accuracy in the MBR high scoring group, 78.0% accuracy for the mid scoring group, to a high of 79.2% accuracy in the low scoring MBR group (Figure 2.6). Hence, the comparison of mean values shown by the histograms (highest to lowest percent accuracy: MBR low, MBR high, MBR full, traditional, MBR mid) differs from the comparison of median values shown by the boxplots (highest to lowest percent accuracy: traditional and MBR low scorers, MBR mid scorers, MBR full group, and MBR high scorers).

Results of classification accuracy by cloud type for traditional and MBR subgroups show the MBR low group outperformed the traditional group by over 8% and the MBR mid and MBR low groups by over 10% and 11% respectively.

For mid-level clouds, the MBR group scored higher than the traditional group, with MBR low, middle, and high groups scoring approximately 10%, 8%, and 7% increases in classification accuracy, respectively (Figure 2.7b). For low-level clouds the MBR high scoring group out performed the traditional group by over 4% while the MBR low and mid groups performed less well than the traditional group (Figure 2.7c). For contrails the traditional group scored 3% to 5% better than all three MBR groups (Figure 2.7d). Results from the Mann-Whitney U test indicate that none of these differences are statistically significant (Table 2.5).

Table 2.5 Results from statistical analysis using the Mann Whitney U non-parametric test of rank sum values across independent samples by cloud type.

Category	Cloud Type	Z value	P value
Traditional / MBR	High clouds	-0.22	0.83
	Mid-level clouds	-0.77	0.44
	Low clouds	-0.10	0.92
	Contrails	1.1	0.28
Traditional / MBR Low	High clouds	-1.1	0.28
	Mid-level clouds	-1.5	0.12
	Low clouds	1.8	0.08
	Contrails	0.0	1.0
Traditional / MBR Mid	High clouds	-0.22	0.83
	Mid-level clouds	-0.77	0.44
	Low clouds	-0.10	0.92
	Contrails	0.65	0.51
Traditional / MBR High	High clouds	-0.22	0.83
	Mid-level clouds	-1.5	0.12
	Low clouds	-0.73	0.46
	Contrails	1.1	0.28

DISCUSSION

This project was developed based on the hypothesis that constructing concept maps would invoke model-based reasoning, improving students' theoretical understanding of atmospheric processes and subsequently improve student cloud classification accuracy. All statistical results indicate that there is no significant difference between cloud classification accuracy of students who participated in the concept mapping intervention and those who did not. However, because of the very small sample size the absence of statistical significance may not indicate that there was no impact from the intervention. The statistics used a non-parametric test that compared median values, which clearly showed no increases in the MBR group. However, histograms of mean values were suggestive of improvements in classification accuracy in the MBR group.

Qualitative analysis of the concept maps in comparison with classification accuracy suggested an alternative hypothesis. Students in the high scoring MBR group were able to effectively utilize concept maps to convey their understanding of their training (Figure 2.8). A few students added visual examples, such as sketches and images of cloud types, into their concept maps. This detail helped fortify their understanding gained from the training and helped with classifying clouds more accurately. An unexpected result was that the low scoring MBR group also showed improvement in classification accuracy. These participants did not create a concept map but rather supplied a written response to the questions (Figure 2.9). These types of responses scored low as concept maps, but they do reflect the participant's ability to constructively study, analyze and reflect on the training even if they did not familiarize themselves with how to construct concept maps and chose to reply in written form. Given this, the MBR training appears to demonstrate the possibility of yielding better data quality when participants engage in an additional step beyond traditional online training that requires them to externally represent their understanding of cloud classification concepts, whether represented through concept maps or textual descriptions. Both the high and low scoring concept mapping groups demonstrated more effort in understanding and learning comprehension compared to the mid scoring concept mapping group, who produced minimalistic and erroneous concept maps. This is consistent with Durak & Topcu (2023), Akpan et al. (2022), and McNeal et al. (2018), who found that increased levels of any activity that required students to externally manipulate new information resulted in increased learning outcomes. Utilizing a modified rubric, or perhaps creating one specifically for scoring the completeness and accuracy of external representations regardless of what form they take could better reflect participant understanding of the training.

CONCLUSION

MBR techniques using external representations show promise for improving the quality of cloud classification data collected by citizen scientists, albeit observation scores between the two training groups were not statistically significant. A top scoring MBR group who constructed concept maps performed better in classifying clouds against the traditionally trained group. A low scoring MBR group, whose responses were in written form rather than concept maps, also performed well. These two groups reflect improved participant comprehension of the cloud classification training, compared with participants who did not construct any external representations or who constructed incomplete or inaccurate external representations. Further investigation on MBR applications in improving data quality related to cloud classification is warranted.

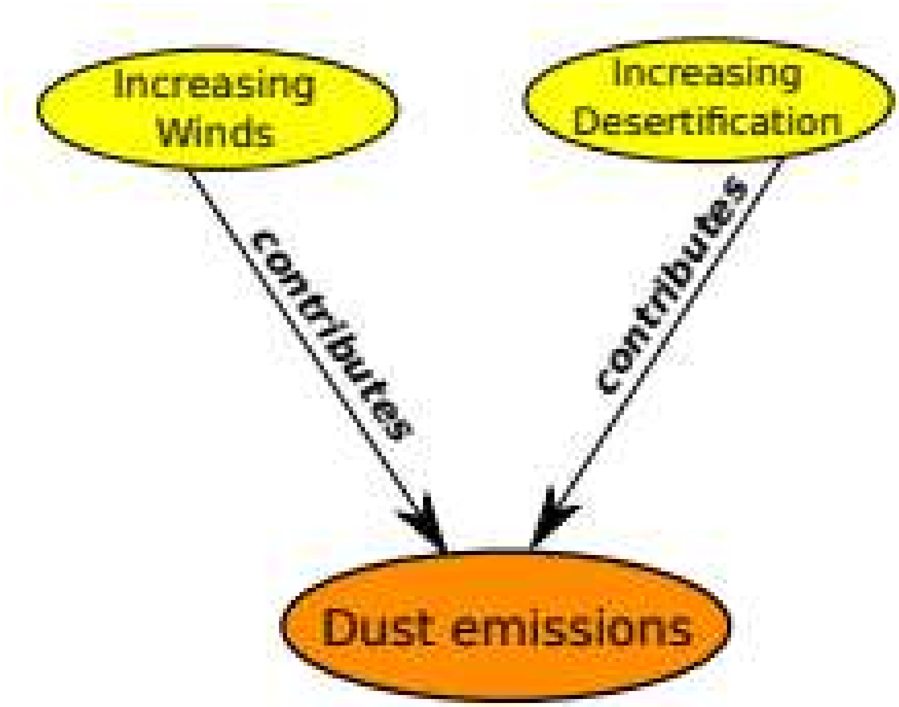
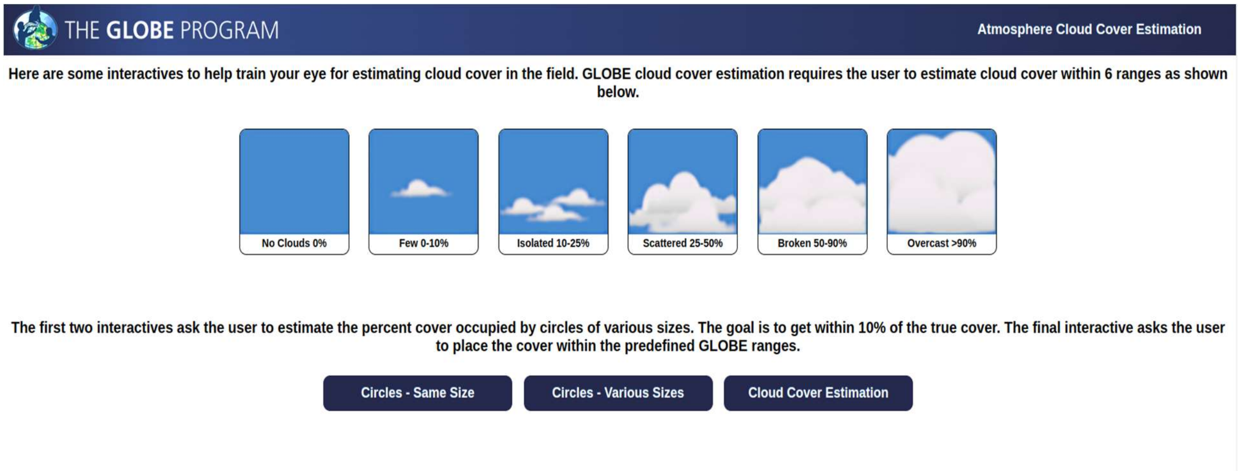


Figure 2.1: Example of a very simple concept map that links three key concepts (shown in ovals) through labeled directed arrows. Each link represents a propositional statement e.g., Increasing Winds contributes to Dust Emissions. Concept maps represent the creator's internal mental models in a way that may be evaluated for correctness and completeness by an instructor.


THE GLOBE PROGRAM

Atmosphere Cloud Cover Estimation

Here are some interactives to help train your eye for estimating cloud cover in the field. GLOBE cloud cover estimation requires the user to estimate cloud cover within 6 ranges as shown below.

The first two interactives ask the user to estimate the percent cover occupied by circles of various sizes. The goal is to get within 10% of the true cover. The final interactive asks the user to place the cover within the predefined GLOBE ranges.

Figure 2.2: Example of online training modules provided by GLOBE (https://observer.globe.gov)

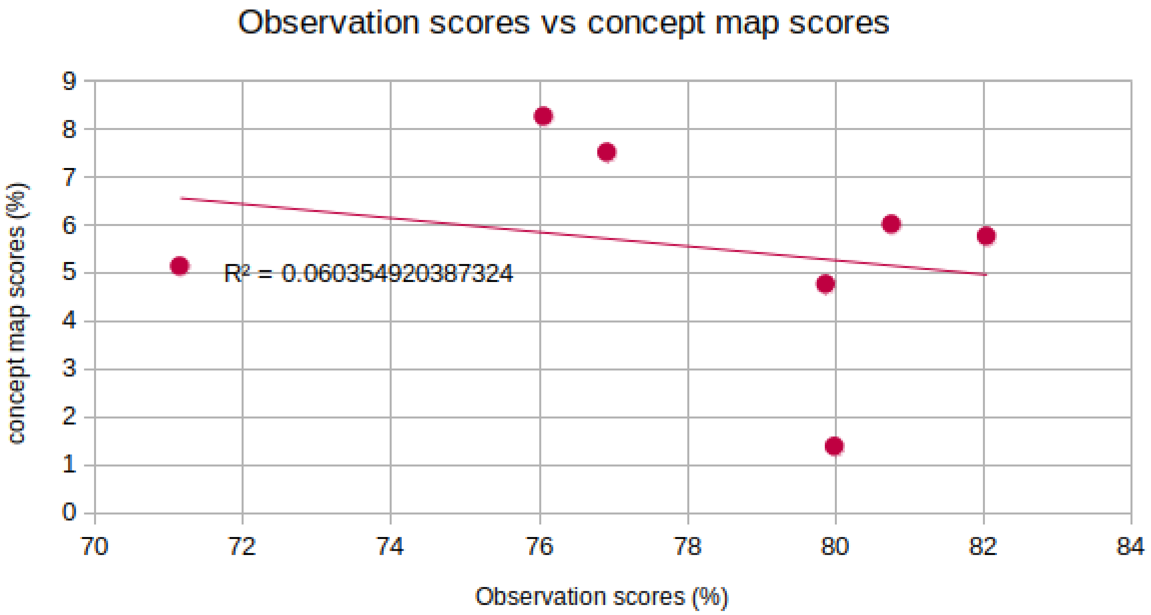


Figure 2.3: Scatterplot of participants' mean concept map score vs. their mean classification accuracy score, showing low correlation between the two. Only participants who completed all eight MBR training modules are included.

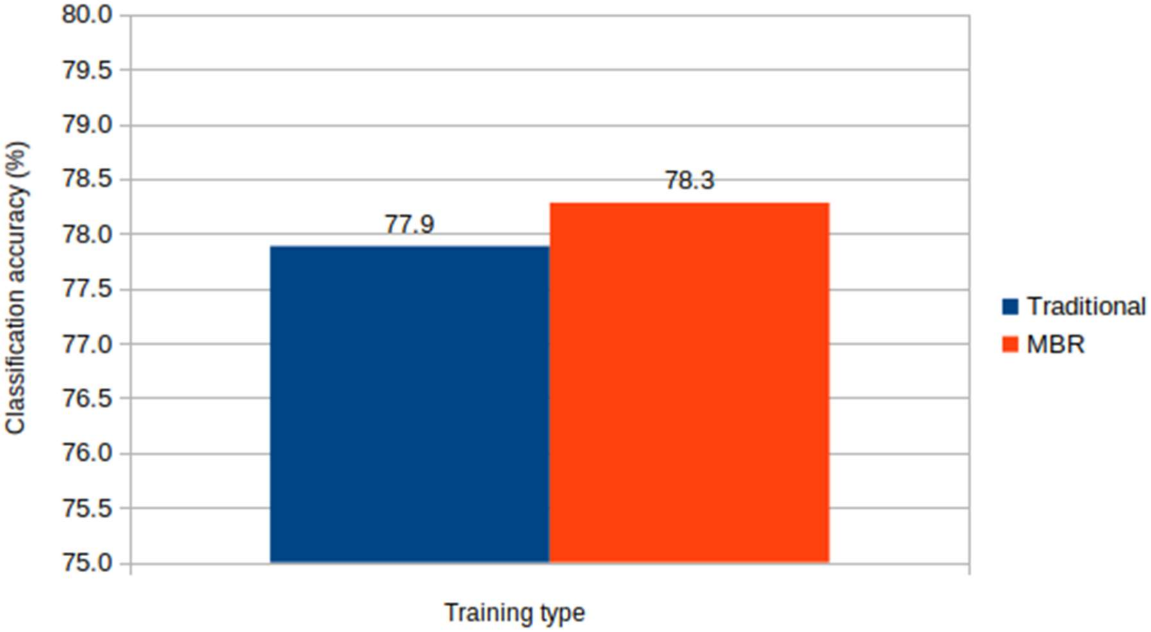


Figure 2.4: Mean cloud classification accuracy for participants trained using traditional online curricula (blue) vs. those trained using traditional curricula plus model based reasoning tasks with concept maps (orange). This shows MBR group with a 0.4%

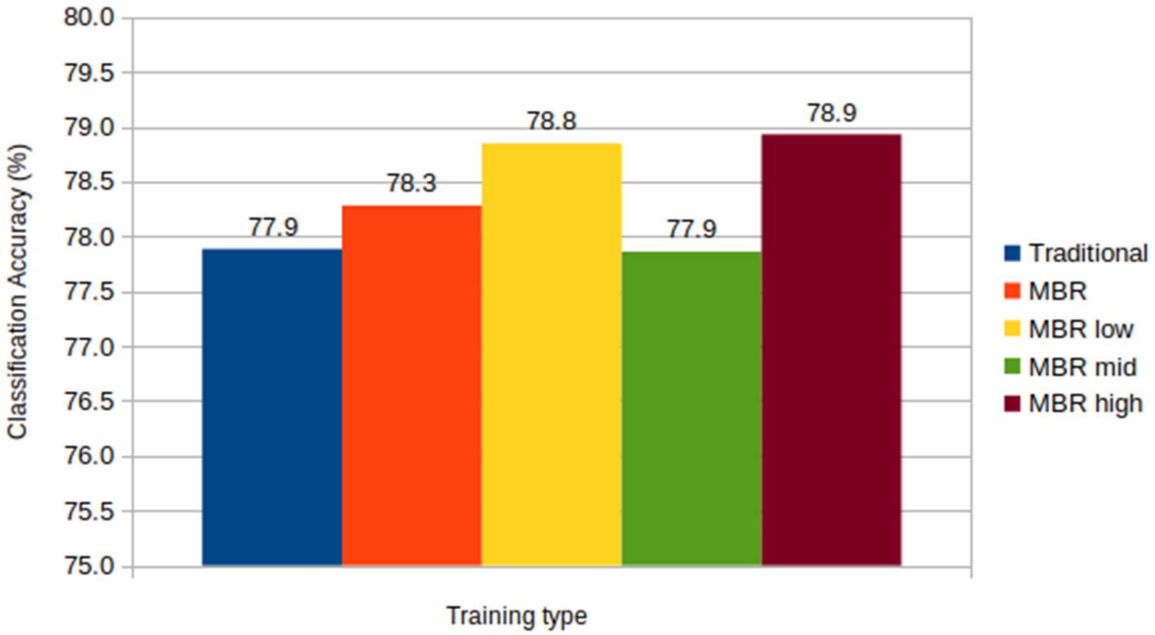


Figure 2.5: Breakdown of traditional vs MBR trained participants. Traditional (blue), MBR trained (orange), low MBR scorers (yellow), middle MBR scorers (green) and high MBR scorers (maroon). Low and high scorers performed better than the traditional group.

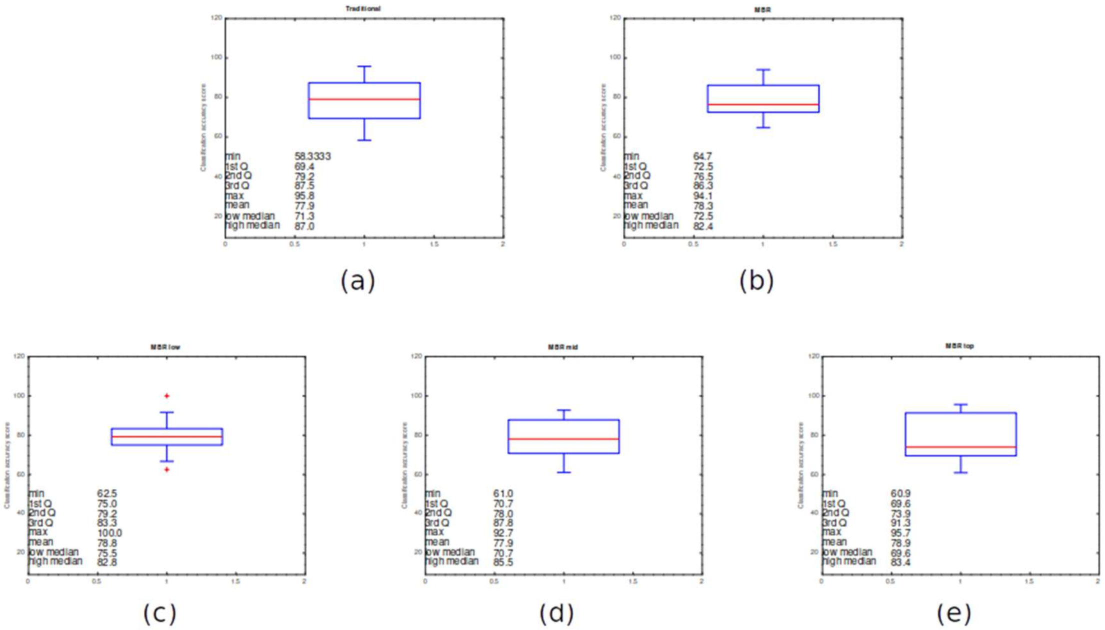


Figure 2.6: Boxplots of classification accuracy data from (a) traditional training, (b) MBR training, (c) MBR low scorers, (d) MBR mid scorers, and (e) MBR high scorers. The classification accuracy scale ranged from 0 to 100 percentage points. The red lines represent the median value. Half of the data fall within the blue box. The lowest and highest quartiles are represented by the blue lines. The red dots in c are outliers.

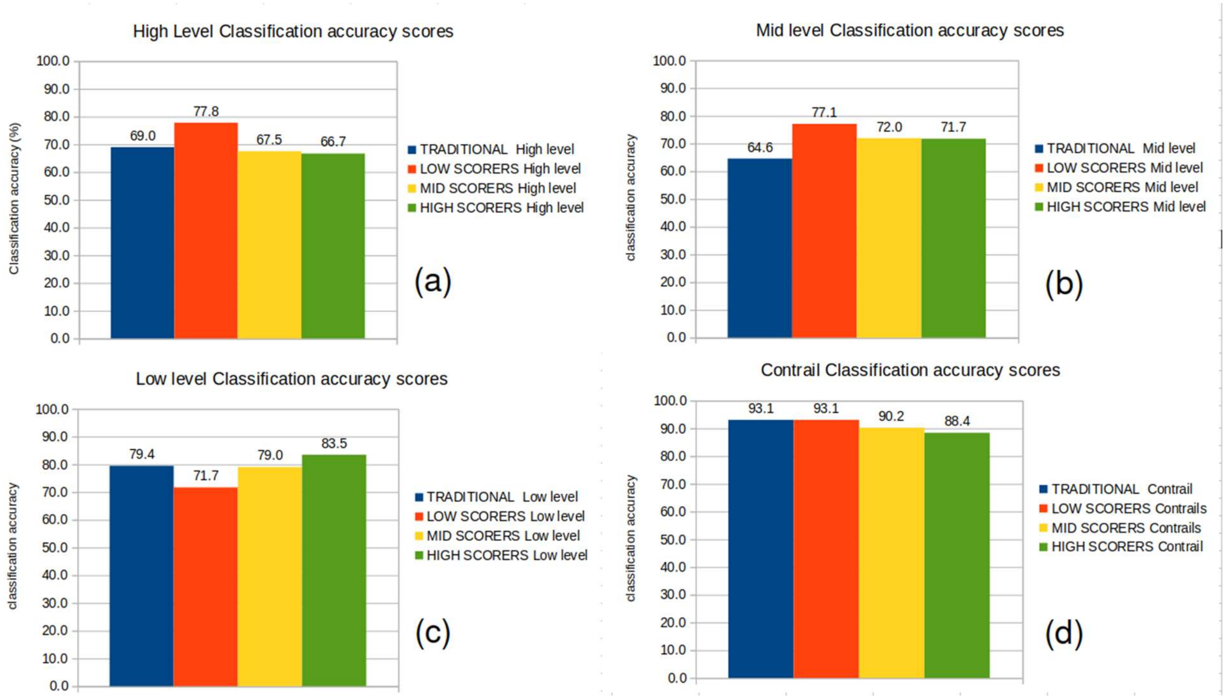


Figure 2.7: Mean cloud classification accuracy by cloud type, for participants trained using traditional online curricula vs. those trained using traditional curricula plus model based reasoning (MBR) tasks with concept maps, subdivided into three groups based on concept map mean scores per participant. A) High level cloud classification accuracy; B) Mid level cloud classification accuracy; C) Low level cloud classification accuracy; and D) Contrail classification accuracy.

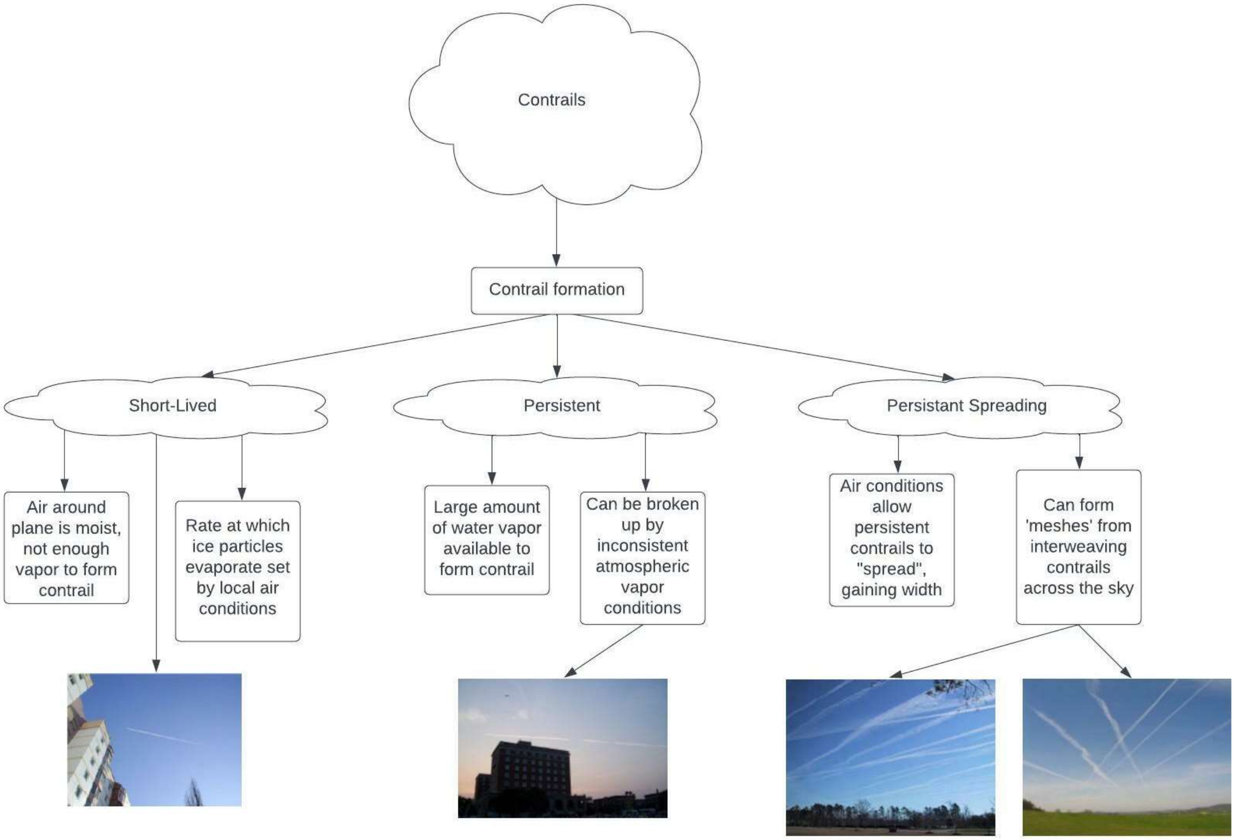
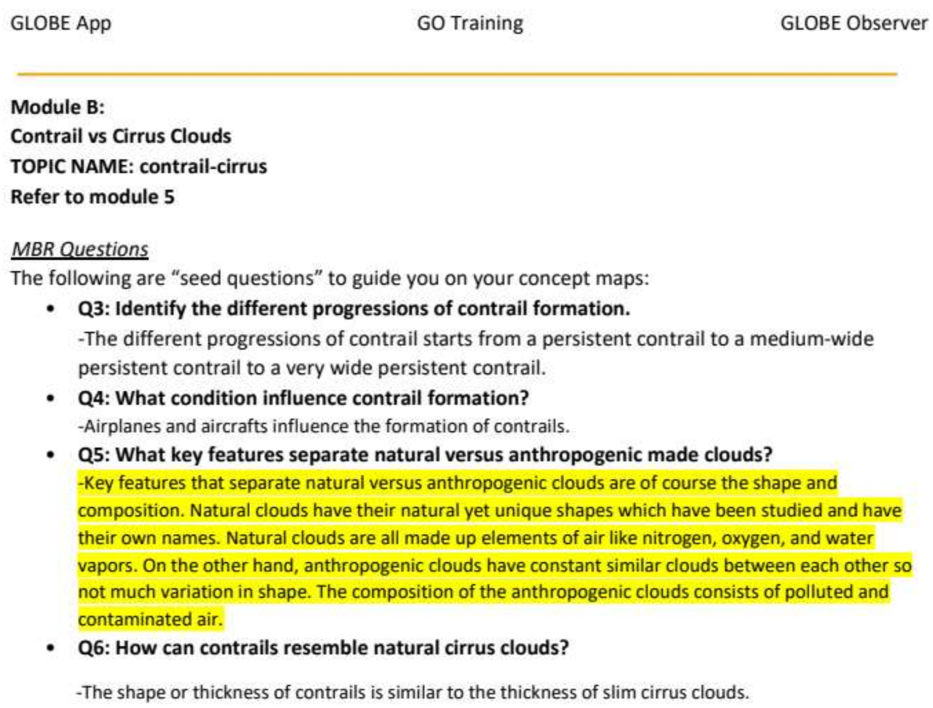

Contrails
Contrail formation
Short-Lived
Persistent
Persistant Spreading
Air around plane is moist, not enough vapor to form contrail
Rate at which ice particles evaporate set by local air conditions
Large amount of water vapor available to form contrail
Can be broken up by inconsistent atmospheric vapor conditions
Air conditions allow persistent contrails to "spread", gaining width
Can form 'meshes' from interweaving contrails across the sky

Figure 2.8: Example of a high scoring concept map. Concepts are comprehensive and links correct, demonstrating good comprehension of cloud classification concepts. Photographs were used to illustrate cloud types.



GLOBE App GO Training GLOBE Observer

Module B:
Contrail vs Cirrus Clouds
TOPIC NAME: contrail-cirrus
Refer to module 5

MBR Questions

The following are "seed questions" to guide you on your concept maps:

- **Q3: Identify the different progressions of contrail formation.**
  -The different progressions of contrail starts from a persistent contrail to a medium-wide persistent contrail to a very wide persistent contrail.
- **Q4: What condition influence contrail formation?**
  -Airplanes and aircrafts influence the formation of contrails.
- **Q5: What key features separate natural versus anthropogenic made clouds?**
  -Key features that separate natural versus anthropogenic clouds are of course the shape and composition. Natural clouds have their natural yet unique shapes which have been studied and have their own names. Natural clouds are all made up elements of air like nitrogen, oxygen, and water vapors. On the other hand, anthropogenic clouds have constant similar clouds between each other so not much variation in shape. The composition of the anthropogenic clouds consists of polluted and contaminated air.
- **Q6: How can contrails resemble natural cirrus clouds?**
  -The shape or thickness of contrails is similar to the thickness of slim cirrus clouds.

Figure 2.9. Example of a low scoring concept map. The participant did not utilize the concept map format, opting for a textual format. The written response (highlighted) demonstrates good comprehension of cloud classification concepts.

NEW MBR/RUBRIC ASSESSMENT

INTRODUCTION

Citizen science has grown in popularity, and with it new opportunities for gathering data for various scientific investigations. A problem that has emerged in data collected by citizen scientists is lower data quality (Amos et al, 2020; Aye et al., 2019; Dodson, 2019; Mitchell et al 2017; Fritz et al, 2017). A common solution to this problem is to collect sufficient citizen science data to perform statistical analysis to identify reliable data (Aye et al., 2019; Dodson, 2019).

New investigative work aimed to address the data quality issue by improving citizen science training that might lead to higher data quality. This research was carried out using data collected by citizen scientists through the NASA GLOBE Observer (GO) application. The first stage of this work was to identify the three most common errors committed by citizen scientists in the GLOBE network. Five hundred datasets from the GLOBE database were randomly selected and analyzed for the top three issues impacting data quality (Olgin and Pennington 2023a, unpublished manuscript):

1) Contrail Confusion – participants confusing natural, high-level cloud formations with anthropogenic cloud formations (i.e. aircraft exhaust);
2) Cumulus Conundrum – issues with differentiating cloud type based on cumulus cloud characteristics among high, mid and low level cloud formations;
3) Coexisting Clouds – multiple cloud types prove confusing in properly classifying cloud formations.

A follow up study investigated the impact of incorporating model-based reasoning (MBR) activities into existing online training provided by GLOBE on cloud classification data quality. MBR activities incorporate visual representation of mental models through diagramming or construction of other graphics and has been demonstrated to improve students' conceptual understanding of complex scientific topics (Kessler et al, 2022; Ubben and Bitzenbauer, 2022; Ifenthaler & Seel, 2013; Jones et al., 2011). The second study incorporated student construction of concept maps around the three cloud classification issues (Olgin and Pennington 2023b, unpublished manuscript). The results were suggestive of a positive impact from the MBR training, yet there was no statistically significant relationship between scores on concept maps and classification accuracy. Additionally, qualitative analysis of the concept maps indicated that some students articulated their correct understanding of the concepts in the form of textual narrative rather than producing a concept map. Although they scored low on the concept maps, their classification accuracy was higher than expected. One of the concluding remarks from that study was that a modified or custom-made rubric was needed to better assess completeness and accuracy of the student's understanding, and the impact of that on classification accuracy.

Prior research has demonstrated the importance of applying the proper rubric to effectively assess learning goals, outcomes and benchmarks (Tractenberg, 2021; Baker et al, 2020; Brookhart, 2018; Janssen et al, 2015). Case studies in the medical sciences (Lo and Wang, 2022), language (Baker et al, 2020) and higher education in general (Pandero and Jonsson, 2022; Baker et al, 2020; Arribas et al, 2017; Brookhard, 2018; Cargas et al, 2017; Dawson, 2017; Sasiprabo et al, 2017) have shown that ineffective rubrics are sometimes unknowningly selected or that proper rubrics may be ineffectively applied. Therefore, to properly assess learning outcomes, it is important to select an appropriate assessment rubric and implement it effectively. Adopting a custom-made rubric to address specific learning outcomes is most effective, as has been reported by Tractenberg (2021), Baker (2020), and Chan and Ho (2019).

The goal of this investigation is to: 1) develop a new rubric applicable to either concept maps or narrative text that better assesses participant comprehension of cloud classification; and 2) test the rubric to see if it better reflects the accuracy of corresponding cloud classifications. The study used the same cloud data, student products, and a similar analytical approach as the prior MBR study. For clarity in distinguishing between the methods and findings of the two studies, this article will refer to the participant products as external representations (EXREP) – including both concept maps and textual narratives – or as concept maps alone (CMAP).

METHODOLOGY

The prior study collected data from two training groups: one traditionally trained through existing online GLOBE training modules. The GLOBE training materials are offered through the GLOBE Observer app and online and include instruction for classifying cloud types, distribution and quantity. The other group using an external representation (EXREP) training scheme where participants used the same traditional online training modules and in addition, created external representations around the three identified classification issues (Contrail Confusion, Cumulus Conundrum, Coexisting Clouds). For each of the three issues, seed questions were developed to target specific learning outcomes and prompt participant generation of relevant external representations (Table 3.1). The traditional and EXREP curricula are summarized in Table 3.2.

Table 3.1 List of three type error classifications gathered from observational data.

Error types	1. Contrail Confusion	2. Cumulus Conundrum	3. Coexisting Clouds
Seed prompts for EXREP concept maps	Q1. Identify the different progression of contrail formation. Q2. What conditions influence contrail formation? Q3. What key features separate natural versus anthropogenic clouds? Q4. How can contrails resemble natural cirrus clouds?	Q5: What are the key characteristics that differentiate altocumulus, cirrocumulus, and cumulus/altocumulus clouds? Q6: What atmospheric and ground conditions are unique to these forms of cumulus cloud formations?	Q7: Identify key conditions that are different between obscurations and overcast skies. Q8: What are at least some conditions that might overlap between obscuration and overcast conditions?

Table 3.2 A comparative summary of traditional versus EXREP training

Traditional training	Model Based Reasoning (EXREP)
1 Participants perform online training through modules for identifying clouds types	4 Participants perform online training through modules for identifying clouds types
2 Assessments are conducted for reflection and auto-feedback from training modules	5 Assessments are conducted for reflection and auto-feedback from training modules
	6 Participants create external representations that address questions in table 3.1.
3 Participants conduct outdoor observations in the field and submit results to NASA's GLOBE database through app	7 Participants conduct outdoor observations in the field and submit results to NASA's GLOBE database through app

Two cohorts of study participants, both in the summer of 2022, were established to test the efficacy of traditional training accompanied with EXREP training when compared to traditional training alone. Both cohorts were freshman/sophomore level students from the El Paso Community College (EPCC) registered for an introductory atmospheric science course that included a module conducting citizen science with the GLOBE Observer app. Human subjects research oversight and approval was obtained from the EPCC Institutional Review Board (1699560-1). Participation in the research was voluntary, consensual, and anonymous.

Each cohort was separated into two groups – those receiving traditional training alone and those receiving additional EXREP training (Table 3.3). The eight in the EXREP group represent those who completed all EXREP training modules.

Table 3.3 Summer 2022 Training Cohorts

Group type	Number of active participants	Number of observations
Traditional	11	80
EXREP	7	56

Each group conducted observations in the field using the Clouds Protocol on the GLOBE Observer app. The Clouds observation protocol specifies the following observations: 1) sky obscurations (fog, heavy rain or snow, sand, spray, smoke, dust, haze or volcanic ash); 2) categorization of any clouds into one of four altitudinal categories and classification of the cloud type within each altitudinal category — very high-level airplane contrails, high-level (cirrus, cirrocumulus, or cirrostratus), mid-level (altostratus or altocumulus), or low-level (fog, nimbostratus, cumulonimbus, stratus, cumulus, or stratocumulus); 3) atmospheric conditions such as an estimate of percent cloud cover and visual opacity (opaque, translucent, or transparent); and 4) surface conditions, including identification of snow/ice, standing water, muddy, dry ground, leaves on trees, and raining/snowing. Multiple cloud types at the same or different levels can be indicated by the participant. The app includes functionality for taking six photos at the time of each observation (north, south, east, west, zenith and nadir photos). All observations were then submitted to the GLOBE cloud database by the participant. Each location where the observation was taken was cataloged with a site ID by GLOBE that participants recorded and submitted to the investigator.

Data associated with each site ID was downloaded from GLOBE by the investigator for analysis. In addition to the observations above, downloaded data included information about the user, the site (latitude, longitude, and elevation), and the date observations were made.

Photos were used to assess cloud classification accuracy for each set of data. A count of the number of correct cloud classifications was recorded for each cloud type, for each cohort and group. Cohorts were combined and accuracy compared between the traditional and EXREP groups using basic statistics, boxplots and histograms.

The EXREP group's external representations were then assessed for completeness and accuracy using a modified rubric, created by the instructor (figure 3.1). This rubric incorporated both analytical and developmental approaches toward assessing concept maps and similar responses that participants issued. The two-part rubric was designed to 1) quantitatively identify the number of propositions created effectively, either using concept maps or written form, and 2) assess the overall quality of the response to determine completeness, accuracy and effectiveness in relaying comprehension of the training. A score ranging from 0 to 1 was given to the number of relatable propositions in response to the prompt given during the EXREP training. A score ranging from 0 to 4 was given to assess the overall quality of the narrative response through the use of either concept maps or written form. The sum of these two parts is then the final score given to that participant's response. The highest scored attained would be a five; the lowest a zero. Those scores were compared with the corresponding observation accuracy using a scatterplot.

The percentage of correct classifications was calculated for the two groups. The EXREP group was subdivided by grouping students with the highest (> 80%), middle (79% - 76%), and lowest (< 76%) external representation scores. The percentage of correct classifications was calculated for the three subset groups. Boxplots were constructed for all five groups. Histograms were created comparing the accuracy of the two unsubdivided groups (traditional and EXREP) and comparing the traditionally trained group with the three subdivided EXREP groups (high, middle, and low scoring concept maps). All five groups were further subdivided by cloud type: high-, mid-, and low-level clouds or contrails. Histograms were constructed comparing classification accuracy by cloud type.

An additional analysis was conducted using the ratio of cloud classification accuracy and EXREP scores compared with cloud classification accuracy. A ratio value of 1.0 or higher indicates the participant gained sufficient expertise from the EXREP training to score well on classification of clouds; a score lower than 1.0 could indicate insufficient expertise was gained from the training. The exploratory statistical analysis was repeated using the ratio value rather than the rubric scores.

A Mann-Whitney U test was used to quantify the statistical significance of differences across independent groups. This is a non-parametric test appropriate for non-normally distributed small samples.

RESULTS

A comparison between external representation scores and classification accuracy was conducted, indicating there is no significant correlation between these data with a calculated R^2 of 0.003

(Figure 3.2). Mean classification accuracy of traditional and EXREP training groups were 77.9% and 78.3% respectively, with the EXREP group scoring approximately 0.4 percentage points higher than the traditional group (Figure 3.3). Based on a breakdown of EXREP types, low scorers to high scorers, the EXREP low scorers underperformed compared to the traditional group by approximately 2.0 percentage points (Figure 3.4). Mid scorers from the EXREP group performed almost 1.0 percentage point better than the traditional group, with the top EXREP scorers outperforming the traditional by 1.5 percentage points. Combining the mid and top scores and comparing with the traditional group shows an almost 1.1% increase in accuracy in cloud classifications.

Results of the Mann Whitney U test indicate that in comparison with the traditional group, there was no significant difference in classification accuracy for any of the EXREP groups (Table 3.4).

Table 3.4 Results from statistical analysis using the Mann Whitney U non-parametric test of rank sum values across independent samples.

Groups	Z value	P value
Traditional / EXREP	-0.08	0.94
Traditional / EXREP Low	-0.08	0.94
Traditional / EXREP Mid	-0.21	0.84
Traditional / EXREP High	-0.28	0.78

Boxplots indicate the median accuracy value of the traditionally trained group was higher than the EXREP full group by 2.7 percentage points (79.2% and 76.5%, respectively; Figure 3.5). Median values for the EXREP subgroups were 75.0% accuracy in the EXREP high scoring group, 79.2% accuracy for the mid scoring group, to a high of 71.4% accuracy in the low scoring EXREP group (Figure 3.6). Hence, the comparison of mean values shown by the histograms (highest to lowest percent accuracy: EXREP high, EXREP mid, EXREP full, traditional and EXREP low) differs from the comparison of median values shown by the boxplots (highest to lowest percent accuracy: EXREP mid, traditional, EXREP full, EXREP high and EXREP low scorers).

Results of classification accuracy by cloud type for traditional and EXREP subgroups show the EXREP low and EXREP high group underperformed the traditional group by over 7% and 4% respectively (Figure 3.6a). The EXREP mid group outperformed all groups, outperforming the traditional group by over 8%.

For mid-level clouds, the EXREP mid and high groups scored higher than the traditional group by over 12% and 5% respectively, with EXREP low underperforming the traditional group by over 9% (Figure 6b). For low-level clouds both the EXREP low and high scoring groups outperformed the traditional group by over 6% while the EXREP low underperformed compared to the traditional group by over 7% (Figure 3.6c). For contrails the traditional group scored over 7% better when compared to the EXREP low and high groups, while the EXREP mid group matched the traditional group with 93.1% (Figure 6d). Results from the Mann-Whitney U test indicate that none of the differences are statistically significant at the 0.05 P-value threshold (Table 3.5). However, only the traditional/EXREP mid group shows a slight tendency toward statistical significance (Table 3.5).

Table 3.5 Results from statistical analysis using the Mann Whitney U non-parametric test of rank sum values across independent samples by cloud type based on the modified rubric.

Category	Cloud Type	Z value	P value
Traditional / EXREP	High clouds Mid-level clouds Low clouds Contrails	-0.22 -0.77 -0.10 1.1	0.83 0.44 0.92 0.28
Traditional / EXREP Low	High clouds Mid-level clouds Low clouds Contrails	0.22 0.0 -1.4 0.65	0.83 1.0 0.17 0.51
Traditional / EXREP Mid	High clouds Mid-level clouds Low clouds Contrails	-1.09 -1.55 1.78 0.0	0.28 0.12 0.08 1.0
Traditional / EXREP High	High clouds Mid-level clouds Low clouds Contrails	-0.22 -0.77 -0.10 1.5	0.83 0.43 0.91 0.13

An additional comparison was made taking the ratio between observation scores and EXREP training scores (OBS/EXREP). A ratio value of 1.0 or above would indicate the participant gained sufficient knowledge from the EXREP training to score equally or better than in classifying clouds; a ratio value below 1.0 would indicate the training was insufficient for the participant to yield successful cloud classification. Figure 7 shows a scatterplot of the OBS/EXREP versus observation scores with minimal correlation of $R^2 = 0.32$. Mean classification accuracy between the traditional and EXREP training groups (figure 3.8) remain the same, with the EXREP low group on par with the traditional group; the EXREP mid group underperforming the traditional by 0.6%; EXREP top outperforming the traditional group by 2.6%.

Results of the Mann Whitney U test indicate that in comparison with the traditional group, there was no significant difference in classification accuracy for any of the EXREP groups (Table 3.6).

Table 3.6 Results from statistical analysis using the Mann Whitney U non-parametric test of rank sum values across independent samples.

Groups	Z value	P value
Traditional / EXREP	-0.08	0.94
Traditional / EXREP Low	-0.08	0.94
Traditional / EXREP Mid	0.17	0.86
Traditional / EXREP High	-0.28	0.78

Boxplots for the median values for the EXREP subgroups were 76.9.0% accuracy in the EXREP high scoring group, 77.3% accuracy for the mid scoring group, to 81.2% accuracy in the low scoring EXREP group (Figure 9). Hence, the comparison of mean values shown by the histograms (highest to lowest percent accuracy: EXREP high, EXREP full, traditional and EXREP mid, and EXREP mid – figure 8). Differs from the comparison of median values shown by the boxplots (highest to lowest: EXREP low, traditional, EXREP mid, EXREP high, and EXREP full).

Results of classification accuracy by cloud type for traditional and EXREP subgroups show the EXREP mid and EXREP high group outperformed the traditional group by over 0.7% and 3% respectively (Figure 3.10a). The EXREP low group underperformed the traditional group by at least 6.5%. For mid-level clouds, all EXREP groups outperformed the traditional group; EXREP low by 7.3%; EXREP mid by 1.3%; EXREP high by 8.5% (Figure 3.10b). For low-level clouds both the EXREP mid and high scoring groups underperformed the traditional group by over 2.1%, with the EXREP low group outperforming the traditional group by 3.1% (Figure 3.10c). For contrails the traditional group scored over 4% better when compared to all EXREP subgroups (Figure 3.10d). Results from the Mann-Whitney U test indicate that only the traditional/EXREP mid group show no statistical significance (Table 3.7).

Table 3.7 Results from statistical analysis using the Mann Whitney U non-parametric test of rank sum values across independent samples by cloud type.

Category	Cloud Type	Z value	P value
Traditional / EXREP	High clouds	-0.22	0.83
	Mid-level clouds	-0.77	0.44
	Low clouds	-0.10	0.92
	Contrails	1.1	0.28
Traditional / EXREP Low	High clouds	0.65	0.51
	Mid-level clouds	-0.77	0.44
	Low clouds	-0.31	0.75
	Contrails	-1.09	0.28
Traditional / EXREP Mid	High clouds	-0.22	0.83
	Mid-level clouds	-0.77	0.44
	Low clouds	0.52	0.60
	Contrails	1.1	0.28
Traditional / EXREP High	High clouds	-1.09	0.28
	Mid-level clouds	-0.77	0.43
	Low clouds	-0.73	0.46
	Contrails	0.22	0.83

DISCUSSION

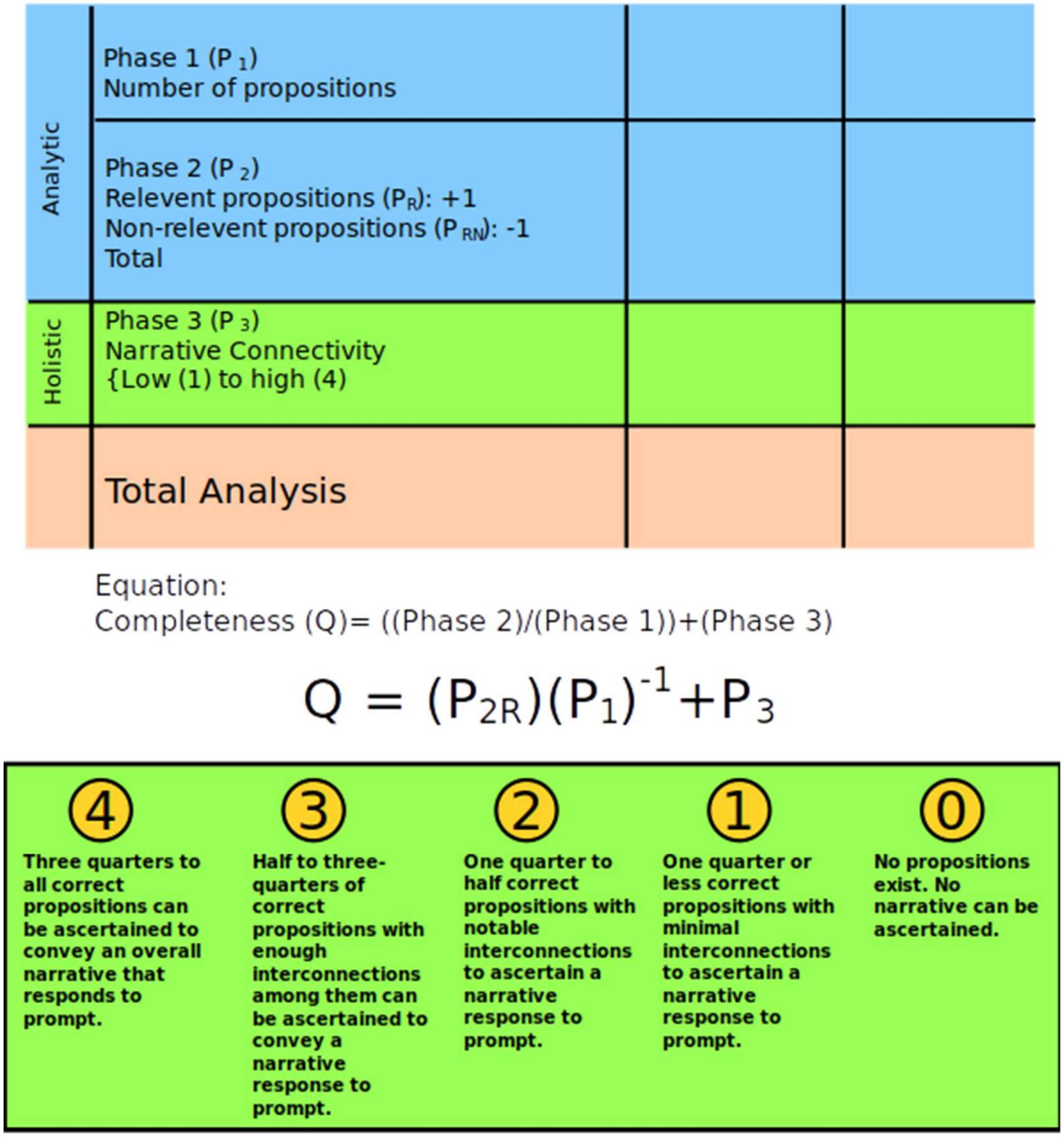
This study was developed to test the efficacy of a modified rubric to better assess the performance of the EXREP group compared with the traditionally trained group. The previous study on CMAP efficacy for improving cloud classification accuracy encountered participants not adhering to the concept mapping approach during the training assignments. Therefore, a modified rubric was created (figure 3.1) that incorporated analytical and developmental formats to account for both connections participants made with learned concepts and their overall comprehension of the training material. This rubric could be applied to either concept map or narrative styled responses. This approach would more effectively highlight the efficacy of additional training using external representations.

The result of this new approach was that redistribution of EXREP subgroups into more predictable outcomes and an incremental improvement of classification accuracy scores with increasing EXREP scores (figure 3.4). Results show that the EXREP low group underperformed when compared to the traditional group, and EXREP mid and high scoring groups outperformed traditional by 0.5% and 1.0% respectively. This assessment is reflective of the modified rubric identifying more accurate participant comprehension of the training materials. Though the Mann-Whitney U test does not show statistical significance of this data, there was an increase in statistical significance, specifically when analyzing the traditional/EXREP mid groups, when compared to the previous study using an established CMAP rubric.

Though there is not a high level of correlation, the $R^2 = 0.31$ of the cloud classification/EXREP score ratio shows a higher correlation than just EXREP – classification accuracy scores ($R^2 < 0.01$). Further study on this approach with an increased sample size is still needed.

CONCLUSION

Participants that responded to the EXREP training did overall perform better than those traditionally trained in cloud classification. This was assessed more effectively through a modified rubric that took into account the different responses made by the participants and highlighted more effectively their comprehension of the training material. The lower scoring EXREP group underperformed compared to the traditional group, whereas the EXREP mid and high scoring groups outperformed the traditional group. Additional investigation is needed to further improve the efficacy of EXREP training when applied to improving data quality.



Analytic	Phase 1 (P_1) Number of propositions		
	Phase 2 (P_2) Relevent propositions (P_R): +1 Non-relevent propositions (P_{RN}): -1 Total		
Holistic	Phase 3 (P_3) Narrative Connectivity {Low (1) to high (4)		
	Total Analysis		

Equation:
Completeness (Q)= ((Phase 2)/(Phase 1))+(Phase 3)

$$Q = (P_{2R})(P_1)^{-1}+P_3$$

4	3	2	1	0
Three quarters to all correct propositions can be ascertained to convey an overall narrative that responds to prompt.	Half to three-quarters of correct propositions with enough interconnections among them can be ascertained to convey a narrative response to prompt.	One quarter to half correct propositions with notable interconnections to ascertain a narrative response to prompt.	One quarter or less correct propositions with minimal interconnections to ascertain a narrative response to prompt.	No propositions exist. No narrative can be ascertained.

Figure 3.1: Rubric used to assess EXREP responses to the EXREP questions. The equation used to determine the overall score: The sum of the number of relevant (or correct) prepositions over the total number of propositions identified and the overall narrative.

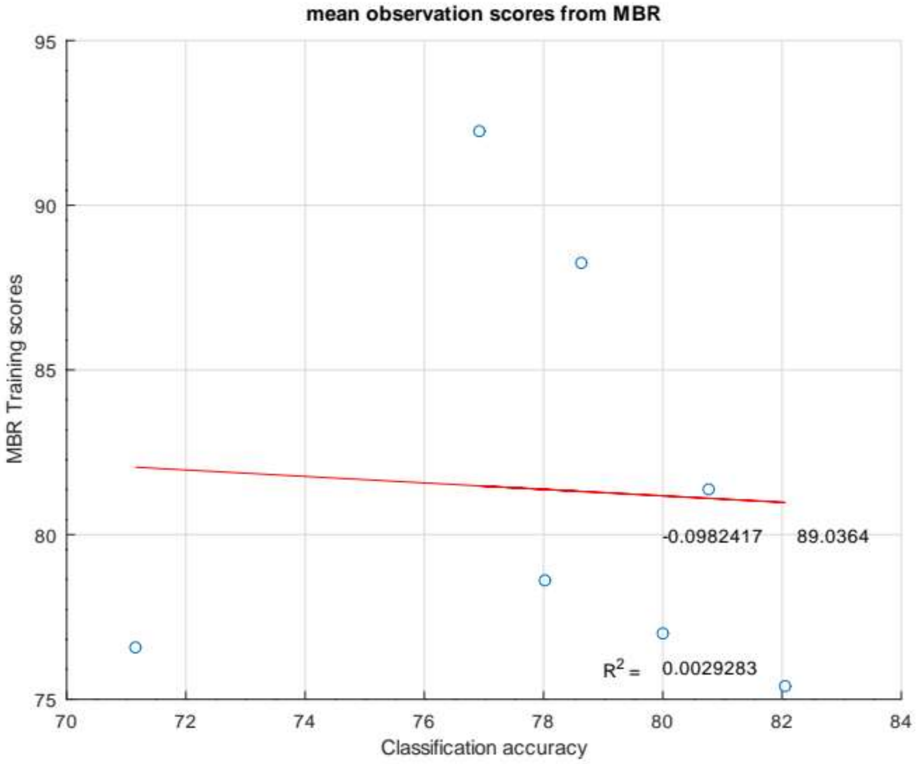


Figure 3.2: Scatterplot of EXREP scores vs. observation scores based on participants responses. An R2 of < 0.01 indicates no correlation among the data.

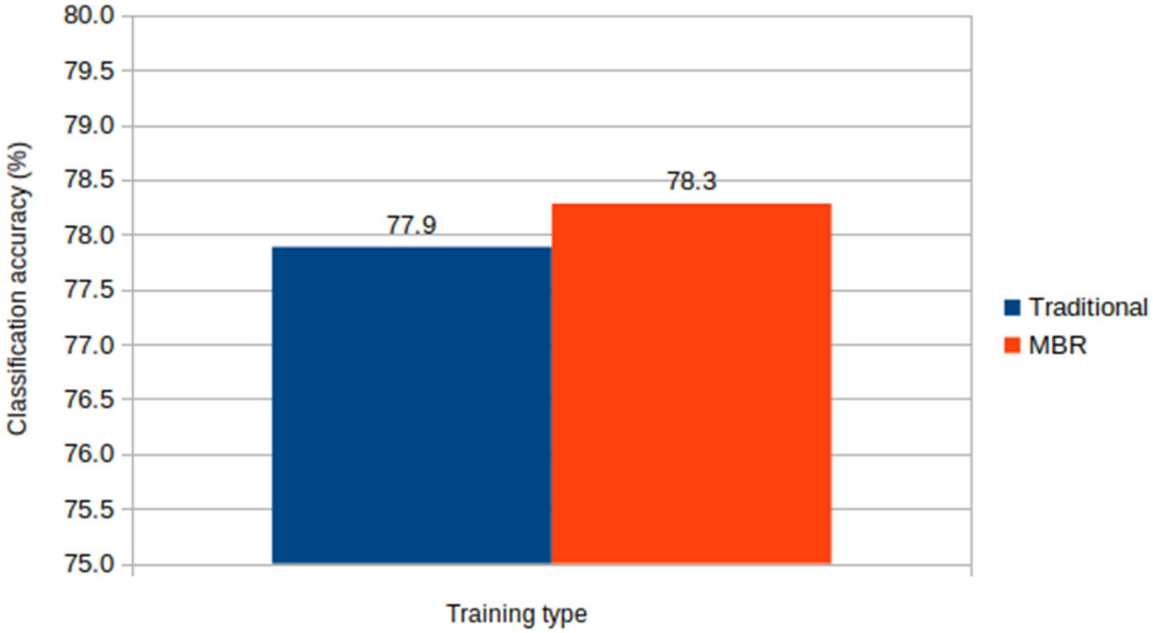


Figure 3.3: Classification accuracy results of the two main training groups: traditional and EXREP; EXREP outperformed traditional training by 0.4%.

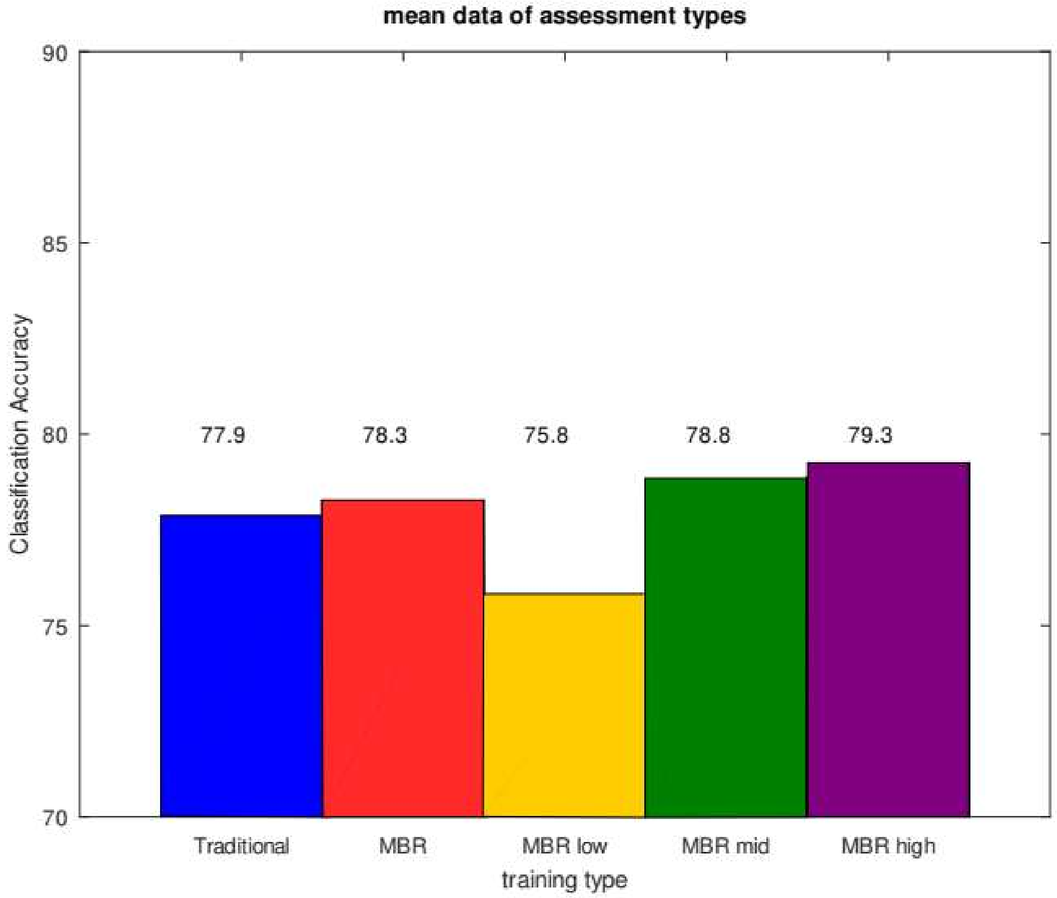


Figure 3.4: Cloud classification accuracy results based on traditional training, EXREP training and EXREP training subgroups. The EXREP low group underperformed compared to the traditional group, EXREP and other EXREP subgroups. Results based on the assessment using the modified rubric.

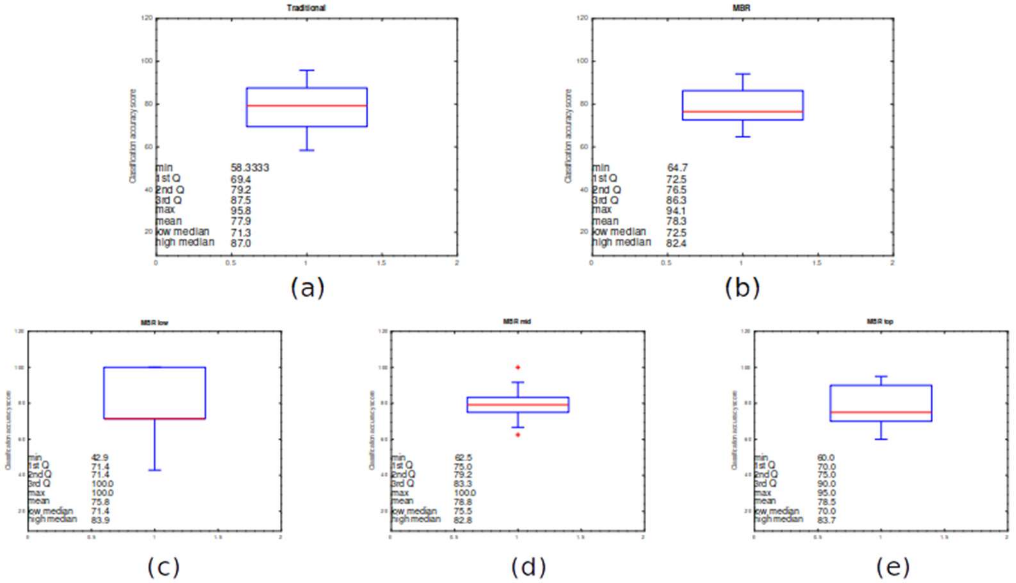


Figure 3.5: Boxplots of classification accuracy data from (a) traditional training, (b) EXREP training, (c) EXREP low scorers, (d) EXREP mid scorers, and (e) EXREP high scorers. The classification accuracy scale ranged from 0 to 100 percentage points. The red lines represent the median value. Half of the data fall within the blue box. The lowest an highest quartiles are represented by the blue lines. The red dots in c are outliers.

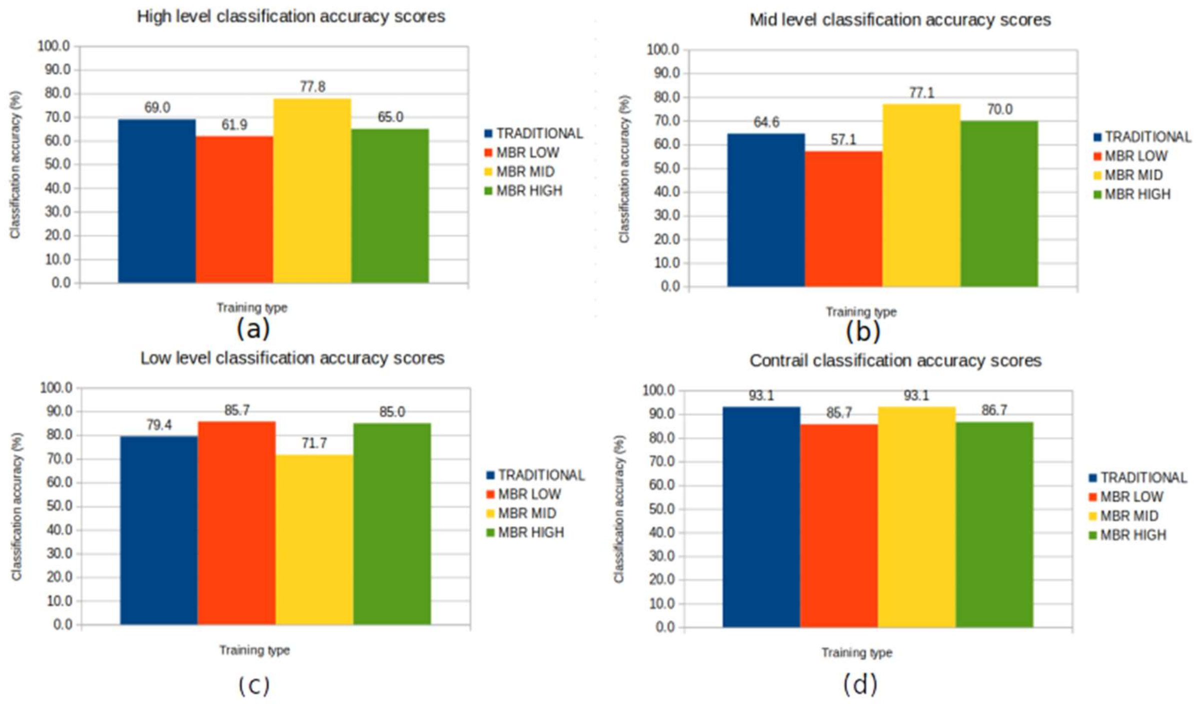


Figure 3.6: Mean cloud classification accuracy by cloud type, for participants trained using traditional online curricula vs. those trained using traditional curricula plus model based reasoning (EXREP) tasks with concept maps, subdivided into three groups based on concept map mean scores per participant. A) High level cloud classification accuracy; B) Mid level cloud classification accuracy; C) Low level cloud classification accuracy; and D) Contrail classification accuracy. Results based on the modified rubric for assessing EXREP responses.

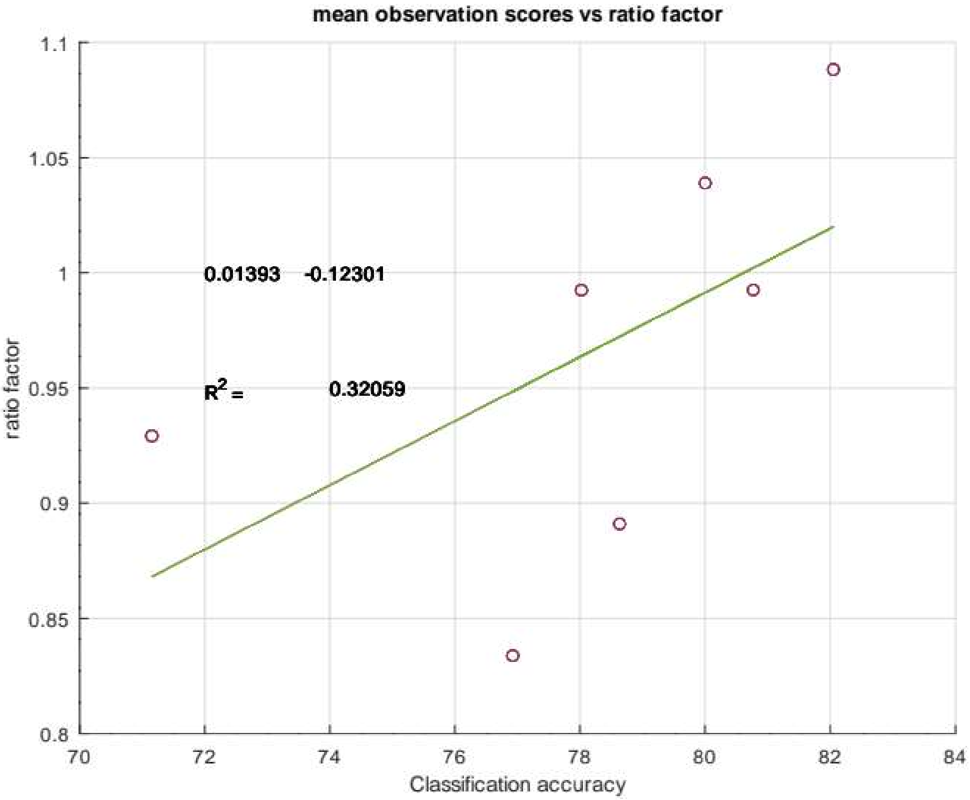


Figure 3.7: Scatterplot of the ratio of cloud classification accuracy and EXREP scores. R2 is higher than previous scatterplot analysis of just cloud classifications (i.e. observations) and EXREP results.

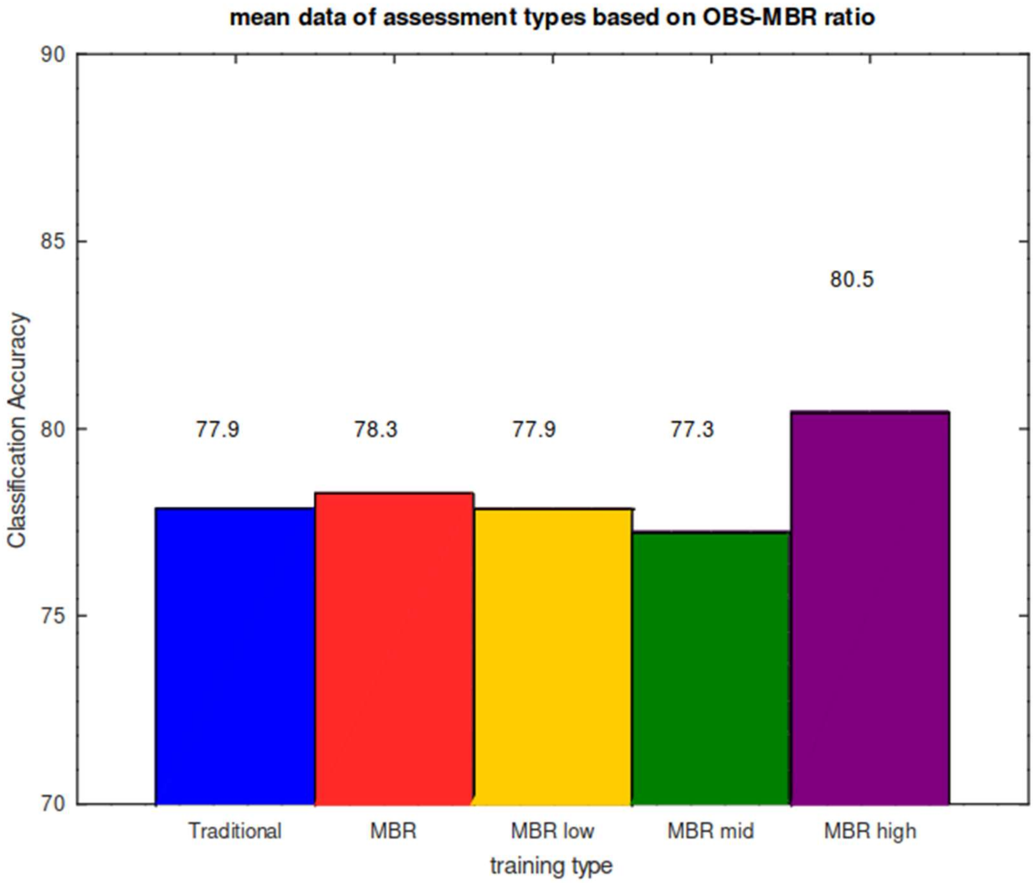


Figure 3.8: Bar plot results of classification accuracy versus training types. The OBS/EXREP ratio was used to categorize participant observation scores. Here EXREP low scores are on par with traditional results, where EXREP mid scores are below par with traditional scores, and EXREP high scores outperforming the traditional group by over 2.5% .

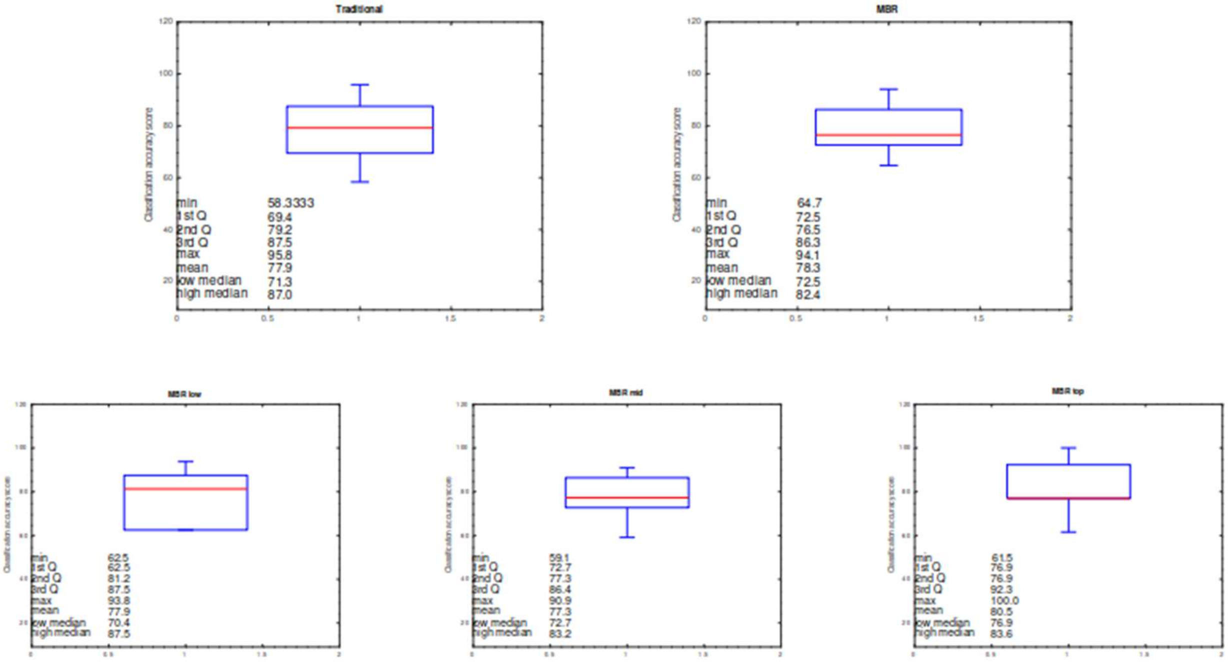


Figure 3.9: Boxplots of classification accuracy data, based on the OBS/EXREP ratio analysis, from (a) traditional training, (b) EXREP training, (c) EXREP low scorers, (d) EXREP mid scorers, and (e) EXREP high scorers. The classification accuracy scale ranged from 0 to 100 percentage points. The red lines represent the median value. Half of the data fall within the blue box. The lowest and highest quartiles are represented by the blue lines. The red dots in c are outliers.

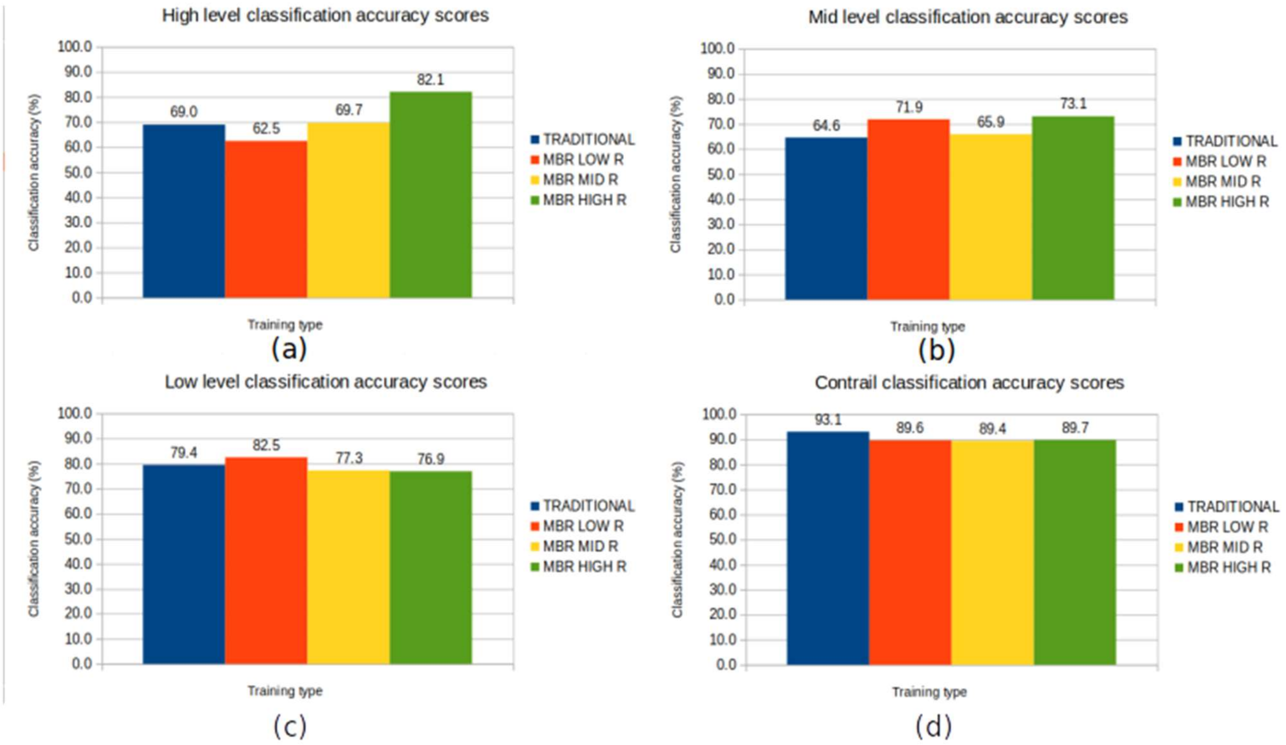


Figure 3.10: Mean cloud classification accuracy by cloud type, for participants trained using traditional online curricula vs. those trained using traditional curricula plus model based reasoning (EXREP) tasks with concept maps, subdivided into three groups based on concept map mean scores per participant. A) High level cloud classification accuracy; B) Mid level cloud classification accuracy; C) Low level cloud classification accuracy; and D) Contrail classification accuracy. Results based on the modified rubric for assessing EXREP responses as well as the OBS/EXREP ratio analysis.

CONCLUSIONS

This comprehensive investigation on data quality from Citizen science participants demonstrated the need for more robust training regiments. MBR techniques, as demonstrated in this work, shows promise in addressing this problem. Identification of key issues with data quality specifically can also help in developing more accurate MBR regiments. The first stage of this investigation went into lengths to identify at least three main errors that were demonstrated by Citizen scientists worldwide. Those were 1) Contrail Confusion – being able to decipher difference between high level clouds and anthropogenic-sourced emission; 2) Cumulus Conundrum – determining the differences between cumulus characteristics between high, mid and low level clouds; and 3) Coexisting Clouds – where all level type clouds could coexist simultaneously making it difficult do separate the differences. What was common among three errors was that determining elevation among these cloud types proved difficult, such as the high and mid level clouds, and separating key differences among the cloud types hampered overall cloud classification.

These errors were then utilized to develop MBR training in an effort to improve data quality with participants. The traditional and MBR training were introduced to them, with the MBR training demonstrating minimal improvement in cloud classification. However the MBR responses were not representative to the actual engagement of participants in the MBR training, as some of them responded outside the intended use of concept maps; an unexpected independence of the participant in this investigation. Therefore the rubric used to assess MBR effectiveness was inadequate, and a custom made one that took into account this response variation was constructed.

The new rubric, which incorporated both holistic and analytical properties, more effectively addressed participant engagement and responses. It allowed for a better qualitative analysis as it allowed the investigator to better follow the thought patterns and logic of trying to comprehend the training. The rubric also allowed to focus more on those that constructed concept maps with greater detail and effort. Results of the use of this new rubric showed that the EXREP groups did performed better overall. Adopting this approach in developing appropriate rubrics for effective and accurate assessments is continually being addressed in other areas in academia (Pandero and Jonsson, 2022; Baker et al, 2020; Arribas et al, 2017; Brookhard, 2018; Cargas et al, 2017; Dawson, 2017; Sasiprabo et al, 2017).

Throughout this investigation, participants were engaging with the training and appeared to improve overall in their data collection. Both training types provided them with a means for more effective engagement and opportunity for increased comprehension of the material. Future study in this area should consist of a model-based reasoning approach that adopts other types and means of responding to training, as well as develop a customized rubric that is specific to that training.

REFERENCES

Aceves-Bueno, E., Adeleye, A. S., Feraud, M., Huang, Y., Tao, M., Yang, Y., & Anderson, S. E. (2017). The Accuracy of Citizen Science Data: A Quantitative Review. *The Bulletin of the Ecological Society of America*, *98*(4), 278–290. https://doi.org/10.1002/bes2.1336

Amos, H. M., Starke, M. J., Rogerson, T. M., Colón Robles, M., Andersen, T., Boger, R., Campbell, B. A., Low, R. D., Nelson, P., Overoye, D., Taylor, J. E., Weaver, K. L., Ferrell, T. M., Kohl, H., & Schwerin, T. G. (2020). GLOBE Observer Data: 2016–2019. *Earth and Space Science*, *7*(8). https://doi.org/10.1029/2020EA001175

Annis, A., & Nardi, F. (2021). GFPLAIN and Multi-Source Data Assimilation Modeling: Conceptualization of a Flood Forecasting Framework Supported by Hydrogeomorphic Floodplain Rapid Mapping. In *Hydrology* (Vol. 8, Issue 4). https://doi.org/10.3390/hydrology8040143

Archer, M. O., Hartinger, M. D., Redmon, R., Angelopoulos, V., & Walsh, B. M. (2018). First Results From Sonification and Exploratory Citizen Science of Magnetospheric ULF Waves: Long-Lasting Decreasing-Frequency Poloidal Field Line Resonances Following Geomagnetic Storms. *Space Weather*, *16*(11), 1753–1769. https://doi.org/10.1029/2018SW001988

Ardon-Dryer, K., Gill, T. E., & Tong, D. Q. (2023). When a Dust Storm Is Not a Dust Storm: Reliability of Dust Records From the Storm Events Database and Implications for Geohealth Applications. *GeoHealth*, *7*(1). https://doi.org/10.1029/2022GH000699

Arribas, E., Ramirez-Vazquez, R., Escobar, I., Gonzalez-Rubio, J., Belendez, A., & Barrera, J. (2019). Development of a laboratory practice for physics introductory courses using a rubric for evaluation by competences. *Journal of Physics: Conference Series*, *1287*(1). https://doi.org/10.1088/1742-6596/1287/1/012025

Arthurs, L. A. (2019). Undergraduate geoscience education research: Evolution of an emerging field of discipline-based education research. *Journal of Research in Science Teaching*, *56*(2), 118–140. https://doi.org/10.1002/tea.21471

Auerbach, J., Barthelmess, E. L., Cavalier, D., Cooper, C. B., Fenyk, H., Haklay, M., Hulbert, J. M., Kyba, C. C. M., Larson, L. R., Lewandowski, E., & Shanley, L. (2019). The problem with delineating narrow criteria for citizen science. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(31), 15336–15337. https://doi.org/10.1073/pnas.1909278116

Ault, T. W., Czajkowski, K. P., Benko, T., Coss, J., Struble, J., Spongberg, A., Templin, M., & Gross, C. (2006). Validation of the MODIS snow product and cloud mask using student and

NWS cooperative station observations in the Lower Great Lakes Region. *Remote Sensing of Environment*, *105*(4), 341–353. https://doi.org/10.1016/j.rse.2006.07.004

Avard, M. M., & Clark, B. K. (2001). Globe in preservice and inservice teacher education. *Journal of Geoscience Education*, *49*(5), 461–466. https://doi.org/10.5408/1089-9995-49.5.461

Aye, K. M., Schwamb, M. E., Portyankina, G., Hansen, C. J., McMaster, A., Miller, G. R. M., Carstensen, B., Snyder, C., Parrish, M., Lynn, S., Mai, C., Miller, D., Simpson, R. J., & Smith, A. M. (2019). Planet Four: Probing springtime winds on Mars by mapping the southern polar CO2 jet deposits. *Icarus*, *319*(August 2018), 558–598. https://doi.org/10.1016/j.icarus.2018.08.018

Baker, B. A., Homayounzadeh, M., & Arias, A. (2020). Development of a test taker-oriented rubric: Exploring its usefulness for test preparation and writing development. *Journal of Second Language Writing*, *50*. https://doi.org/10.1016/j.jslw.2020.100771

Balázs, B., Mooney, P., Nováková, E., Bastin, L., & Jokar Arsanjani, J. (2021). Data Quality in Citizen Science. *The Science of Citizen Science*, 139–157. https://doi.org/10.1007/978-3-030-58278-4_8

Barrutia, O., Ruíz-González, A., Villarroel, J. D., & Díez, J. R. (2021). Primary and Secondary Students' Understanding of the Rainfall Phenomenon and Related Water Systems: a Comparative Study of Two Methodological Approaches. *Research in Science Education*, *51*, 823–844. https://doi.org/10.1007/s11165-019-9831-2

Benedetti, Y., Slezak, K., Møller, A. P., Morelli, F., & Tryjanowski, P. (2018). Number of syllables in cuckoo Cuculus canorus calls: A test using a citizen science project. *Scientific Reports*, *8*(1), 1–6. https://doi.org/10.1038/s41598-018-31329-1

Brookhart, S. M. (2018). Appropriate Criteria: Key to Effective Rubrics. In *Frontiers in Education* (Vol. 3). Frontiers Media S.A. https://doi.org/10.3389/feduc.2018.00022

Callaghan, C. T., Poore, A. G. B., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., Rowley, J. J. L., Vergés, A., Wilshire, J. H., & Cornwell, W. K. (2021). Three Frontiers for the Future of Biodiversity Research Using Citizen Science Data. *BioScience*, *71*(1), 55–63. https://doi.org/10.1093/biosci/biaa131

Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B., & Major, R. E. (2019). Improving big citizen science data: Moving beyond haphazard sampling. *PLoS Biology*, *17*(6), 1–11. https://doi.org/10.1371/journal.pbio.3000357

Cargas, S., Williams, S., & Rosenberg, M. (2017). An approach to teaching critical thinking across disciplines using performance tasks with a common rubric. *Thinking Skills and Creativity*, *26*, 24–37. https://doi.org/10.1016/j.tsc.2017.05.005

Cervato, C., Charlevoix, D., & Gold, A. (2018). *Research on Students ' Conceptual Understanding of Environmental , Oceanic , Atmospheric , and Climate Science Content*. 17–34.

Charmaz, K., & Thornberg, R. (2021). The pursuit of quality in grounded theory. *Qualitative Research in Psychology*, *18*(3), 305–327. https://doi.org/10.1080/14780887.2020.1780357

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, *7*, 205031211882292. https://doi.org/10.1177/2050312118822927

Cuzzolino, M. P., Grotzer, T. A., Tutwiler, M. S., & Torres, E. W. (2019). An agentive focus may limit learning about complex causality and systems dynamics: A study of seventh graders' explanations of ecosystems. *Journal of Research in Science Teaching*, *56*(8), 1083–1105. https://doi.org/10.1002/tea.21549

Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment and Evaluation in Higher Education*, *42*(3), 347–360. https://doi.org/10.1080/02602938.2015.1111294

de Sherbinin, A., Bowser, A., Chuang, T. R., Cooper, C., Danielsen, F., Edmunds, R., Elias, P., Faustman, E., Hultquist, C., Mondardini, R., Popescu, I., Shonowo, A., & Sivakumar, K. (2021). The Critical Importance of Citizen Science Data. *Frontiers in Climate*, *3*(March), 1–7. https://doi.org/10.3389/fclim.2021.650760

Deutsch, E. S., Cardille, J. A., Koll-Egyed, T., & Fortin, M. J. (2021). Landsat 8 lake water clarity empirical algorithms: Large-scale calibration and validation using government and citizen science data from across Canada. *Remote Sensing*, *13*(7). https://doi.org/10.3390/rs13071257

Dodson, J. B., Colón Robles, M., Rogerson, T. M., & Taylor, J. E. (2023). Do Citizen Science Intense Observation Periods Increase Data Usability? A Deep Dive of the NASA GLOBE Clouds Data Set With Satellite Comparisons. *Earth and Space Science*, *10*(2), 1–18. https://doi.org/10.1029/2021EA002058

Dodson, J. B., Robles, M. C., Taylor, J. E., Defontes, C. C., & Weaver, K. L. (2019). Eclipse across america: Citizen science observations of the 21 august 2017 total solar eclipse. *Journal of Applied Meteorology and Climatology*, *58*(11), 2363–2385. https://doi.org/10.1175/JAMC-D-18-0297.1

Durak, B., & Topçu, M. S. (2023). Integrating socioscientific issues and model-based learning to decide on a local issue: the white butterfly unit. *Science Activities*, *60*(2), 90–105. https://doi.org/10.1080/00368121.2023.2179967

Edgett, K. S., & Christensen, P. R. (1996). K-12 Education Outreach Program Initiated by a University Research Team for the Mars Global Surveyor Thermal Emission Spectrometer Project. *Journal of Geoscience Education*, *44*(2), 183–188. https://doi.org/10.5408/1089-9995-44.2.183

Eisen, L., & Eisen, R. J. (2021). Benefits and drawbacks of citizen science to complement traditional data gathering approaches for medically important hard ticks (Acari: Ixodidae) in the United States. *Journal of Medical Entomology*, *58*(1), 1–9. https://doi.org/10.1093/jme/tjaa165

Enterkine, J., Campbell, B. A., Kohl, H., Glenn, N. F., Weaver, K., & Overoye, D. (n.d.). *The potential of citizen science data to complement satellite and airborne lidar tree height measurements : lessons from The GLOBE Program OPEN ACCESS The potential of citizen science data to complement satellite and airborne lidar tree height measurement*.

Follett, R., Strezov, V., Peter, M., Diekötter, T., Kremer, K., Roy, H. E., Pocock, M. J. O., Preston, C. D., Roy, D. B., Savage, J., Tweddle, J. C., Robinson, L. D., Aceves-Bueno, E., Adeleye, A. S., Feraud, M., Huang, Y., Tao, M., Yang, Y., Anderson, S. E., … Ehleringer, J. R. (2019). Design and impact of the national workshop for early career geoscience faculty. *Biological Conservation*, *59*(1), 106266. https://doi.org/10.1016/j.chb.2020.106266

Fritz, S., Fonte, C. C., & See, L. (2017). The role of Citizen Science in Earth Observation. *Remote Sensing*, *9*(4). https://doi.org/10.3390/rs9040357

Fritz, S., See, L., Carlson, T., Haklay, M. (Muki), Oliver, J. L., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L. A., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., … West, S. (2019). Citizen science and the United Nations Sustainable Development Goals. *Nature Sustainability*, *2*(10), 922–930. https://doi.org/10.1038/s41893-019-0390-3

Gold, A. U., Pendergast, P. M., Ormand, C. J., Budd, D. A., & Mueller, K. J. (2018). Improving spatial thinking skills among undergraduate geology students through short online training exercises. *International Journal of Science Education*, *40*(18), 2205–2225. https://doi.org/10.1080/09500693.2018.1525621

Graham, E. A., Henderson, S., & Schloss, A. (2011). Using mobile phones to engage citizen scientists in research. *Eos*, *92*(38), 313–315. https://doi.org/10.1029/2011EO380002

Henderson, S., Hatheway, B., Gardiner, L., & Zarlengo, K. (2006). An Early Introduction to Earth System Science through Elementary GLOBE. *Journal of Geoscience Education*, *54*(3), 210–214. https://doi.org/10.5408/1089-9995-54.3.210

Hunter, J., Alabri, A., & van Ingen, C. (2013). Assessing the quality and trustworthiness of citizen science data. *Concurrency and Computation: Practice and Experience*, *25*(4), 454–466. https://doi.org/https://doi.org/10.1002/cpe.2923

Ifenthaler, D., & Seel, N. M. (2013a). Model-based reasoning. *Computers and Education*, *64*, 131–142. https://doi.org/10.1016/j.compedu.2012.11.014

Ifenthaler, D., & Seel, N. M. (2013b). Model-based reasoning. *Computers & Education*, *64*, 131–142. https://doi.org/10.1016/j.compedu.2012.11.014

Johnson, E. T., & McNeal, K. S. (2022). Student perspectives of the spatial thinking components embedded in a topographic map activity using an augmented-reality sandbox. *Journal of Geoscience Education*, *70*(1), 13–24. https://doi.org/10.1080/10899995.2021.1969862

Jollymore, A., Haines, M. J., Satterfield, T., & Johnson, M. S. (2017). Citizen science for water quality monitoring: Data implications of citizen perspectives. *Journal of Environmental Management*, *200*(2017), 456–467. https://doi.org/10.1016/j.jenvman.2017.05.083

Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental models: An interdisciplinary synthesis of theory and methods. *Ecology and Society*, *16*(1). https://doi.org/10.5751/ES-03802-160146

Jones, P. D., & Reid, P. A. (2001). Temperature trends in regions affected by increasing aridity/humidity. *Geophysical Research Letters*, *28*(20), 3919–3922. https://doi.org/10.1029/2001GL013840

Kessler, S. H., Schäfer, M. S., Johann, D., & Rauhut, H. (2022). Mapping mental models of science communication: How academics in Germany, Austria and Switzerland understand and practice science communication. *Public Understanding of Science*, *31*(6), 711–731. https://doi.org/10.1177/09636625211065743

Koffler, S., Barbiéri, C., Ghilardi-Lopes, N. P., Leocadio, J. N., Albertini, B., Francoy, T. M., & Saraiva, A. M. (2021). A buzz for sustainability and conservation: The growing potential of citizen science studies on bees. *Sustainability (Switzerland)*, *13*(2), 1–15. https://doi.org/10.3390/su13020959

Langenkämper, D., Simon-Lledó, E., Hosking, B., Jones, D. O. B., & Nattkemper, T. W. (2019). On the impact of Citizen Science-derived data quality on deep learning based classification in marine images. *PLoS ONE*, *14*(6), 1–16. https://doi.org/10.1371/journal.pone.0218086

Lepenies, R., & Zakari, I. S. (2021). Citizen science for transformative air quality policy in Germany and Niger. *Sustainability (Switzerland)*, *13*(7). https://doi.org/10.3390/su13073973

Levitt, H. M. (2021). Essentials of critical-constructivist grounded theory research. In *Essentials of critical-constructivist grounded theory research.* American Psychological Association. https://doi.org/10.1037/0000231-000

Liu, Z., & Stasko, J. (2010). Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 999–1008. https://doi.org/10.1109/TVCG.2010.177

Low, R., Boger, R., Nelson, P., & Kimura, M. (2021). GLOBE Mosquito Habitat Mapper Citizen Science Data 2017–2020. *GeoHealth*, *5*(10). https://doi.org/10.1029/2021GH000436

Lukyanenko, R., Wiggins, A., & Rosser, H. K. (2019). Citizen Science: An Information Quality Research Frontier. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-019-09915-z

MacDonald, E. A., Donovan, E., Nishimura, Y., Case, N. A., Megan Gillies, D., Gallardo-Lacourt, B., Archer, W. E., Spanswick, E. L., Bourassa, N., Connors, M., Heavner, M., Jackel, B., Kosar, B., Knudsen, D. J., Ratzlaff, C., & Schofield, I. (2018). New science in plain sight: Citizen scientists lead to the discovery of optical structure in the upper atmosphere. *Science Advances*, *4*(3), 16–21. https://doi.org/10.1126/sciadv.aaq0030

McLaughlin, J. A., & Bailey, J. M. (2022). Students need more practice with spatial thinking in geoscience education: a systematic review of the literature. *Studies in Science Education*, *00*(00), 1–58. https://doi.org/10.1080/03057267.2022.2029305

McNeal, P., Ellis, T., & Petcovic, H. (2018). Investigating the foundations of spatial thinking in meteorology. *Journal of Geoscience Education*, *66*(3), 246–257. https://doi.org/10.1080/10899995.2018.1483119

McNeal, P. M., & Petcovic, H. L. (2020). Spatial thinking and fluid Earth science education research. *Journal of Geoscience Education*, *68*(4), 289–301. https://doi.org/10.1080/10899995.2020.1768007

Mitchell, N., Triska, M., Liberatore, A., Ashcroft, L., Weatherill, R., & Longnecker, N. (2017). Benefits and challenges of incorporating citizen science into university education. *PLoS ONE*, *12*(11), 1–15. https://doi.org/10.1371/journal.pone.0186285

Nersessian, N. J. (2009). How do engineering scientists think? Model-based simulation in biomedical engineering research laboratories. *Topics in Cognitive Science*, *1*(4), 730–757. https://doi.org/10.1111/j.1756-8765.2009.01032.x

Newcombe, N., & Shipley, T. F. (2015). Studying Visual and Spatial Reasoning for Design Creativity. *Studying Visual and Spatial Reasoning for Design Creativity*, *November*. https://doi.org/10.1007/978-94-017-9297-4

Newman, G., Graham, J., Crall, A., & Laituri, M. (2011). The art and science of multi-scale citizen science support. *Ecological Informatics*, *6*(3–4), 217–227. https://doi.org/10.1016/j.ecoinf.2011.03.002

Panadero, E., & Jonsson, A. (n.d.). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review.*, *9*.

Peplow, M. (2018). Poisonous politics in the Rust Belt. *Nature*, *559*, 180.

Pernat, N., Kampen, H., Jeschke, J. M., & Werner, D. (2021). Citizen science versus professional data collection: Comparison of approaches to mosquito monitoring in Germany. *Journal of Applied Ecology*, *58*(2), 214–223. https://doi.org/10.1111/1365-2664.13767

Reges, H. W., Doesken, N., Turner, J., Newman, N., Bergantino, A., & Schwalbe, Z. (2016). CoCoRaHS: The evolution and accomplishments of a volunteer rain gauge network. *Bulletin of the American Meteorological Society*, *97*(10), 1831–1846. https://doi.org/10.1175/BAMS-D-14-00213.1

Robles, M. C., Amos, H. M., Dodson, J. B., Bouwman, J., Rogerson, T., Bombosch, A., Farmer, L., Burdick, A., Taylor, J., & Chambers, L. H. (2020). Clouds around the World. *American Meteorological Society*, *February 2020*, 1201–1213.

Sasipraba, T., Kaja Bantha Navas, R., Nandhitha, N. M., Prakash, S., Jayaprabakar, J., Poorna Pushpakala, S., Subbiah, G., Kavipriya, P., Ravi, T., & Arunkumar, G. (2020). Assessment tools and rubrics for evaluating the capstone projects in outcome based education. *Procedia Computer Science*, *172*, 296–301. https://doi.org/10.1016/j.procs.2020.05.047

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654. https://doi.org/10.1002/tea.20311

Sezen-Barrie, A., Henderson, J. A., & Drewes, A. L. (2022). *Spatial and Temporal Dynamics in Climate Change Education Discourse: An Ecolinguistic Perspective BT - Critical Thinking in Biology and Environmental Education: Facing Challenges in a Post-Truth World* (B. Puig & M. P. Jiménez-Aleixandre (Eds.); pp. 189–209). Springer International Publishing. https://doi.org/10.1007/978-3-030-92006-7_11

Smolleck, L. D., Zembal-Saul, C., & Yoder, E. P. (2006). The development and validation of an instrument to measure preservice teachers' self-efficacy in regard to the teaching of science as inquiry. *Journal of Science Teacher Education*, *17*(2), 137–163. https://doi.org/10.1007/s10972-006-9015-6

Stylinski, C. D., Peterman, K., Phillips, T., Linhart, J., & Becker-Klein, R. (2020). Assessing science inquiry skills of citizen science volunteers: a snapshot of the field. *International*

Journal of Science Education, Part B: Communication and Public Engagement, *10*(1), 77–92. https://doi.org/10.1080/21548455.2020.1719288

Sullivan, K. O., & Ed, S. (n.d.). *An Application of Concept Mapping for Instruction and Assessment*. 310.

Tierney, G., Goodell, A., Nolen, S. B., Lee, N., Whitfield, L., & Abbott, R. D. (2020). (Re)Designing for Engagement in a Project-based AP Environmental Science Course. *Journal of Experimental Education*, *88*(1), 72–102. https://doi.org/10.1080/00220973.2018.1535479

Ubben, M. S., & Bitzenbauer, P. (2022). Two Cognitive Dimensions of Students' Mental Models in Science: Fidelity of Gestalt and Functional Fidelity. *Education Sciences*, *12*(3). https://doi.org/10.3390/educsci12030163

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. In *Psychological Bulletin* (Vol. 139, pp. 352–402). American Psychological Association. https://doi.org/10.1037/a0028446

Van Haeften, S., Milic, A., Addison-Smith, B., Butcher, C., & Davies, J. M. (2021). Grass Gazers: Using citizen science as a tool to facilitate practical and online science learning for secondary school students during the COVID-19 lockdown. *Ecology and Evolution*, *11*(8), 3488–3500. https://doi.org/10.1002/ece3.6948

Vann-Sander, S., Clifton, J., & Harvey, E. (2016). Can citizen science work? Perceptions of the role and utility of citizen science in a marine policy and management context. *Marine Policy*, *72*, 82–93. https://doi.org/10.1016/j.marpol.2016.06.026

VITA

John Gilbert Olgin grew up learning about astronomy and passionate to teach others about the wonders of science. He received his Bachelor's Degree in Physics in 2002, Masters in Physics (2006) and geology (2012), and Ph.D. in geological sciences (2023). He has taught both at the University of Texas at El Paso (UTEP) and El Paso Community College (EPCC) in the fields of physics, astronomy and geology. He was a geophysicist at Shell plc, working on a project related to CO2 sequestration. His planetary geology research on *Limits of Enceladus's ice shell thickness from tidally driven tiger stripe shear failure* was published in Geophysical Research Letters in 2011 (https://doi.org/10.1029/2010GL044950). He has conducted numerous education/public outreach activities at both UTEP and EPCC, and works with NASA's GLOBE Program as a GLOBE teacher on public outreach and research projects involving clouds, dust storms and solar eclipses https://www.globe.gov/web/jolgin.2). His is currently pursuing a tenure track teaching position at El Paso Community College and other adventures in science.

John G. Olgin
jolgin@epcc.edu