Proceedings for the
3rd Shaw-IAU Workshop
on Astronomy for Education

What Everybody Should Know about Astronomy Education

12 – 15 October, 2021

Editors: Asmita Bhandare, Giuliana Giobbi, Colm Larkin, Rebecca Sanderson, Eduardo Penteado, Niall Deacon, Gwen Sanderson, and Anna Sippel

Publications of the
IAU Office of Astronomy for Education

Compiled & Edited by:
Asmita Bhandare, Giuliana Giobbi, Colm Larkin, Rebecca Sanderson, Eduardo Penteado, Niall Deacon, Gwen Sanderson, and Anna Sippel.

The following is a summary of the 3rd Shaw-IAU workshop on Astronomy for Education held 12 – 15 October, 2021 as a virtual event. The workshop was organised by the IAU Office of Astronomy for Education. More details can be found on: https://astro4edu.org/shaw-iau/3rd-shaw-iau-workshop/.

The Office of Astronomy for Education (OAE) is hosted by the Haus der Astronomie on the campus of the Max Planck Institute for Astronomy in Heidelberg. The OAE's mission is to support and coordinate astronomy education by astronomy researchers and educators, aimed at primary or secondary schools worldwide. The OAE is an office of the International Astronomical Union, with substantial funding from the Klaus Tschira Foundation and the Carl Zeiss Foundation. The Shaw-IAU Workshops on Astronomy for Education are funded by the Shaw Prize Foundation.

Foreword

3rd Shaw-IAU Workshop on Astronomy for Education

Teaching astronomy takes both solid knowledge of the subject itself as well as educational skills, such as knowing appropriate methods and techniques for teaching. To this, specific sub-fields of astronomy education add their own specialized skill sets: knowing how to operate remote telescopes, for instance, or the ins and outs of daytime observations. Last but not least, there are the skills needed in order to make our teaching fair, equitable, and inclusive.

In practice, most of us who are active in astronomy education have only been taught a subset of those skills in our academic training. Those who come from professional astronomy and have branched out into education and outreach typically have advanced training in astronomy, but not in the relevant areas of pedagogy. Most teachers, on the other hand, have pedagogical training as well as training in the subjects their teach, but often that does not include formal training in astronomy and astronomy education.

If this description includes you, and if in consequence you have ever felt motivated to expand your astronomy education skill set, then this workshop was, and is, meant for you. It is the third in a series organised as a collaborative venture between the Shaw Prize Foundation and the International Astronomical Union, and with 89 talks and 50 posters in a total of 18 sessions, it provides a fairly comprehensive "Astronomy Education 101".

For those who were unable to attend, or did not manage to attend all of the sessions they were interested in, we present these proceedings, and the associated talk videos from the workshop. While they lack the interactivity that the 580 workshop participants enjoyed as they posed their questions to the speakers, or interacted in the chat, we do believe that they are valuable in their own right — and we asked speakers to include in their write-ups helpful pointers to additional resources, so you have the opportunity to delve deeper. If you find these resources useful, and I hope they will be useful to many, please share them widely.

The workshop was made possible by funding from the Shaw Prize Foundation, for which we are very grateful. You can find the names of the individuals and institutions who organised the workshop on p. 6 — a big "Thank you!" to all of you!

For us at the International Astronomical Union's Office of Astronomy for Education (IAU OAE), this is just the start. Helping those who are active in astronomy education to grow their skills, and to become more professional in their activities, is one of our main objectives. Stay in touch if you want to make sure not to miss what is next — from additional events to more resources. On the web, you can find us at http://astro4edu.org, and on that page, you can also find your country's National Astronomy Education Coordinator Team. We are also on Twitter and on Facebook as @astro4edu.

Markus Pössel
Director, IAU Office of Astronomy for Education
Heidelberg, November 16, 2021

Contents

Organizing Committees

Local Organizing Committee:
Asmita Bhandare, Suresh Bhattarai, Niall Deacon, Theo Dimitrakoppoulos, Natalie Fischer, Giuliana Giobbi, Celine Greis, Esther Kolar, Colm Larkin, Tshiamiso Makwela, Carmen Müllerthann, Markus Nielbock, Eduardo Penteado, Markus Pössel, Saeed Salimpour, Gwen Sanderson, Rebecca Sanderson, Anna Sippel

Organizing Committee:
Asmita Bhandare, Silvia Casu, Niall Deacon, Dario del Moro, Samir Dhurde, Giuliana Giobbi, Edward Gomez, Renate Hubele, Tshiamiso Makwela, Sivuyile Manxoyi, Markus Nielbock, Eduardo Penteado, Markus Pössel (Chair), Sara Ricciardi, Rosa M. Ros, Saeed Salimpour, Stefano Sandrelli, Anna Sippel, Aniket Sule, Stefania Varano, Alessandra Zanazzi

Scientific Advisory Committee:
Asmita Bhandare, Silvia Casu, Niall Deacon, Dario del Moro, Samir Dhurde, Urban Eriksson, Michael Fitzgerald, Giuliana Giobbi, Edward Gomez, Andrej Guštin, Marietta Gyulzadyan, Renate Hubele, Fraser Lewis, Carolin Liefke, Tshiamiso Makwela, Hakim Luthfi Malasan, Sivuyile Manxoyi, Surhud More, Markus Nielbock, Arvind Paranjpye, Eduardo Penteado, Markus Pössel, Carmelo Presicce, Travis Rector, Sara Ricciardi, Nayra Rodríguez Eugenio, Gustavo Rojas, Rosa M. Ros, Saeed Salimpour, Stefano Sandrelli, Anna Sippel, Aniket Sule, Rachele Toniolo, Rosa Valiante, Stefania Varano, Sarita Vig, Alessandra Zanazzi, Anita Zanella

In addition to the efforts from the OAE office in Heidelberg, Germany, the following OAE Centers and Node made key contributions to organizing this event:

OAE Center Italy

OAE Center Cyprus

OAE Node Nepal

The OAE Center India was not formally established at the time of this workshop but also made significant contributions.

Evaluation of Resources and Activities

Session organiser: Silvia Casu, INAF - National Institute for Astrophysics, OAE Center Italy, Italy

Session overview

Evaluation is a very important topic when talking about education. Evaluation is a continuous process that critically examines a program; it can improve program design and implementation, assess its achievements and improve upon its effectiveness. It helps teachers and learners to improve teaching and learning processes. Evaluation helps us to make evidence-based decisions.

There are **different types of evaluation**, depending on evaluation purposes. There are evaluation activities that you conduct before you start your work and as you are planning (Front-end evaluation), while a project is in development (Formative evaluation), and at the conclusion of the program implementation (Summative or Impact evaluation).

There are **different types of data** you can collect to perform evaluation: quantitative data (numbers such as simple counts or percentage) and qualitative data (more descriptive in nature). Therefore, there are **different types of methods and tools** one can use to collect and analyse data, to understand *what* is happening and *why*: graphs, closed-ended surveys, checklists, rubrics but also interviews, focus groups, open-ended surveys and more interactive data collection methods. Moreover, we could use mixed methods in order to have a more holistic view.

This session aims to give a general and basic overview of types, methods and tools commonly used in evaluation, together with a list of open resources, to help teachers but also scientists and educators to plan the best assessment of their work.

An Introduction to Evaluation for Astronomy Education Programs

Speaker: Sanlyn Buxner, Planetary Science Institute and University of Arizona, USA

Evaluation plays an important role in developing and revising astronomy education programs and activities as well as understanding the short and long-term impacts on students, teachers, and members of the public. This presentation will give an overview of different types of evaluation including internal and external evaluation as well as needs assessment, formative, and summative evaluation. We will cover the difference between different types of assessments used in evaluation and how these types of data can be used to gain valuable information about your astronomy education activities. We will provide resources for assessment and discuss the difference between astronomy education research and evaluation. Lastly, we will provide resources for getting started doing evaluation.

Talk link: https://youtu.be/vmIo85aXJ3A

Evaluation plays an important role in developing and revising astronomy education programs and activities as well as understanding the short and long-term impacts on students, teachers, and members of the public. In general, assessment is important in helping us, as scientists and educators, make evidence-based decisions. Assessment helps us understand the needs of our audiences and our communities, understand the quality of our products and activities, and understand the alignment of our programs and activities to audience needs. Additionally, assessment helps us gather information to improve our programs and understand their impact.

I encourage everyone to evaluate their programs, big or small. If you do not evaluate your activities and programs, how will you know you have been successful? You use evaluation to figure out what you are doing well and what you can still improve. Lastly, you can use evaluation to convince others to give you time, money, and other resources.

According to the American Evaluation Association (https://www.eval.org/), program evaluation answers questions like: *To what extent does the program achieve its goals? How can it be improved? Should it continue? Are the results worth what the program costs? What are the long-term impacts? A program evaluation has to be designed to be appropriate for the specific program being evaluated.*

There are different types of evaluation that are important in any discussion about evaluation. There are evaluation activities that you conduct before you start your work and as you are planning, activities that you conduct during your work as you implement activities that will inform changes, and activities that you carry out at the conclusion of the program implementation. Below is an overview of the main types of evaluation.

Front end evaluation is conducted at the beginning of the project, often before you apply for funding. This may include a needs assessment where you get input from the audience or community you want or plan to work with. This is important as it tells you the types of things your audience needs or wants and avoids having you produce a program or activity that will not have an appropriate audience. This also helps you build trust and a roadmap about what you want to achieve. You may also consider doing a literature review of either peer reviewed literature, evaluation reports, or other reports that will inform your work. Lastly, reviewing other projects may give you important insight for your new program.

Formative evaluation is conducted while you are implementing a program. This type of evaluation helps you understand the quality of the implementation of your program. If you are running a short event, this may be just asking how things are going and what things can make it better in real time. If you have a multiple-day workshop, this may include some end-of-day questions to improve the experience for the next day. Formative can be an individual activity or for big chunks of your program. The most important aspect is that you want to collect data that is actionable. You also must be willing and able to make changes to your program. Collecting this data and making changes can demonstrate that you are being responsive based on feedback through an iterative feedback cycle.

Summative evaluation is conducted at the end of a project, or a piece of a project, to understand the impacts of the project. This is sometimes referred to as outcomes evaluation or impact evaluation. The overall purpose of summative evaluation is to report on the impact of your program on the intended audience and understand how engaging in your program or activity has had value for the audience. The summative evaluation is often what is shared in a report with your funding agency and can help you answer more long-term impacts about your program. When we think about the actual outcomes we are interested in, it is often helpful to think about the categories of outcomes. These might include participant knowledge, skills, achievement, interests, beliefs, motivations, attitudes, behaviors and choices. A nice framework for thinking about different categories of impact comes from the National Science Foundation Informal Science Education program document (http://www.Informalscience.org/framework-evaluating-impacts-informal-science-education-projects).

When we think about participant knowledge skills, and achievement there are a variety of ways we can measure these. These include using tests, grades on assignments on classes, concept inventories (specific surveys that are designed to get deep insight into students' knowledge about a specific topic), knowledge surveys, as well as interviews and focus group interviews of either participants or their instructors or supervisors. We can also review students' work in class, portfolios of their work or even make observations of performance-based tasks. If we are curious about participants' interests, beliefs, motivations, or attitudes, using surveys of interviews and focus groups will be the easiest way to gather information. If we are interested in looking at behavior and choices, we may choose to conduct follow up surveys or interviews of focus group interviews, or we may choose to look at records of behavior such as school and

employment records. It is important to note that any tool you use should be for the data you want to gather so that it accurately asks what you are interested in. As you are thinking through the types of data you want to collect, it is important to ask, "How will I know that I have achieved my goals?" Once you decide on the types of data you want to collect, you will want to think about a threshold for success? Do you want to make sure everyone has a certain amount of knowledge that there is an increase? These questions both drive what you collect but also how you analyze data.

It is also important to think about whether we want to measure participants' knowledge, interest or motivations at one point in time, if we want to be able to show changes, or if we want to make claims about long-term impact. These types of decisions will tell us if we need to have a single survey (one point in time) or a pre- and post survey to show changes over time. If we want to make claims about changes in participants' knowledge or attitudes, we need to make sure to take a baseline (pre-) before the program so that we can feel confident that there was a change as assessed on the post survey.

There are two broad types of data that we talk about in evaluation. Quantitative data is data to which we assign a numerical score or ask participants to assign a value to something. Examples of quantitative data may include test scores, agreement scales (to what extent do you agree with), basic descriptive statistics (how many people chose this session). Qualitative data is descriptive data that is often in participants' words. Examples of qualitative data include open ended survey responses, essays, observation notes, and interviews. Qualitative data help us make sense of quantitative data and give us more information. It is important to remember that any type of data can be quantified including interview data (e.g. word counts). Each type of data lends itself to different types of questions and in evaluation, I often encourage folks to use both types of data to support different types of evaluation questions. One example of using both might be to conduct a large (1,000) quantitative survey of student knowledge about knowledge of Solar System exploration, interest in STEM, and career plans. Additionally, you might collect qualitative data through interviews of 20 students to better understand their survey responses, get more information about their interest responses (why they rated questions a certain way) and ask more deeply about career plans.

When we think about the difference between evaluation and research, it is important to remember that they are similar in data collection and analysis, but different in their overall purpose. Research is about understanding phenomena and generalizing, and evaluation is about making judgement about the value of your program. There is a nice discussion on https://www.betterevaluation.org/en/blog/framing_the_difference_between_research_and_evaluation.

Evaluation may include both internal and external evaluation. Internal evaluation often involves someone who is close to the project, who works on the project, and who may be invested in the project. Internal evaluators are often less expensive and faster in feedback. External evaluation involves someone who is not associated with the project, who is outside the power structure of the project, who can give a different perspective and honest feedback without a close connection. External evaluation is often more time-consuming and more expensive. External evaluators are often required by funding agencies. The overall take-home message is that evaluation helps you show the value of your work, gives you important feedback to improve and can be done on a small or large scale. Additionally, evaluation can be done by anyone! You can review your own

programs, or help out a colleague. When looking for an evaluator, there are online tools such as the "find an evaluator tool" (https://my.eval.org/find-an-evaluator) of the American Evaluation Association university teaching and learning centers, dedicated evaluation centers at different institutes, your colleagues, other projects who have evaluators, and your network here.

I have put together a list of resources (https://tinyurl.com/4stxyhf2) that include guides to conducting evaluation, guides to preparing evaluation proposals and pieces of evaluation in proposals for funding, links to data collection tools including surveys, interview protocols, observation protocols, and rubrics for assessing work, resources for quantitative and qualitative analysis, sources of evaluation reports, and links to standards for evaluation.

Evaluation Basics: Planning for Improvement and Measuring Impact

Speaker: Amy Grack Nelson, Science Museum of Minnesota, USA

Evaluation is an important tool for improving astronomy education projects and assessing their impact. So how do you go about carrying out an evaluation? Learn about the key components of the evaluation process including identifying the goals of the evaluation, choosing the best data collection method, carrying out the study, analyzing the data, and sharing the results. This session will also share a number of resources to help support you in developing a basic evaluation plan of your own.

Talk link: https://youtu.be/fKPimWQjvIQ

What goes into planning and carrying out an evaluation?: At the most basic level, planning and carrying out an evaluation includes identifying the purpose of the evaluation, selecting data collection methods, identifying the sample, analyzing the data, and reporting results.

Evaluation Purpose: First, we need to identify the purpose of the evaluation. Two typical purposes of evaluation are for improvement (formative evaluation) and to measure impact (summative evaluation).

Formative evaluation takes place while a project is being developed. The purpose of formative evaluation is to identify areas of improvement to help ensure a project is meeting its

intended goals and outcomes. Ideally, formative evaluations are an iterative process where data is gathered, changes are made, and additional data is gathered to see what additional improvements are necessary. Formative evaluation can look at areas such as comprehension, confusion, enjoyment, and ease of use.

Summative evaluation takes place when a project is out of the development phase. The purpose of summative evaluation is to measure the project's impact on intended audiences, or how well the project was able to meet its outcomes. Ideally, results from a summative evaluation can also be used to further improve a project, or inform future projects.

Data Collection Methods: We then need to identify how we are going to gather data to answer our evaluation questions. What methods are going to give us the type of information we need given the structure of the program we are evaluating and the audiences we are gathering data from? Data collection methods can include surveys, interviews, focus groups, interviews, and more interactive data collection methods.

Sample: We also need to think about the sample. Whom are we gathering data from and how will we obtain our sample? You might decide to collect data from everyone participating in a program, or you may want to select or recruit a sample of people.

Data Analysis: Our data collection methods can give us quantitative or qualitative data. We get quantitative data from things like closed-ended survey questions or tallies of observing particular behaviors using a checklist. We get qualitative data from methods like interviews or notes from freeform written observations describing what we are seeing. Often we want to use mixed methods in an evaluation, both quantitative and qualitative, as it gives us a more holistic view of the program.

Quantitative data are numbers, so we might look at simple counts or percentages. If appropriate, we might carry out statistical analyses. We often use graphs to visualize these results in reports and make the findings easier for readers to understand.

Qualitative data are descriptive in nature and can give us deeper insights into the experiences people are having. These data are often analyzed by creating codes that are applied to the data. These codes can be ones that you already have or codes that emerge from responses. After you code the data, you create themes based on those codes and report how frequently those themes appeared in the data. You might also include example quotes from the data in a report to help exemplify the findings.

Reporting: There are various ways you can share your findings. It might be a formal report or a brief or memo where you summarize findings in a more digestible format. Maybe instead of a report you have a meeting with program staff to discuss the results and make sense of the data together. There are a variety of ways you can share the data and it is important to consider the audience when thinking about the best way to share findings with them.

Resources:

The Center for the Advancement of Informal Science Education (CAISE) has a wide variety of resources to help you plan your evaluation on their website `informalscience.org`. Visit their

"Design Evaluation" section, as well as search their collection of thousands of evaluation reports to find examples you can learn from. Below is a list of additional resources that are helpful for designing evaluations of informal science education experiences.

- Center for Advancement of Informal Science Education. (2011). Principal investigator's guide: Managing evaluation in informal STEM education projects. Washington, DC: Author. urlhttps://www.informalscience.org/evaluation/pi-guide

- Feder, M. A., Shouse, A. W., Lewenstein, B., & Bell, P. (Eds.). (2009). Learning Science in Informal Environments: People, Places, and Pursuits. Washington DC: National Academies Press. Free download here http://www.nap.edu/catalog/12190/learning-science-in-informal-environments-people-places-and-pursuits

- Friedman, A. (Ed.). (2008). Framework for evaluating impacts of informal science education projects.

- Report from a National Science Foundation Workshop. From: www.informalscience.org/documents/Eval_Framework.pdf

- Fu, A.C., Kannan, A., & Shavelson, R.J. (Eds.) (2019). Evaluation in informal science, technology, engineering, and mathematics education. New Directions for Evaluation, 161. https://onlinelibrary.wiley.com/toc/1534875x/2019/2019/161

- Pattison, S., Cohn, S., & Kollmann, L. (2014). Team-based inquiry: A practical guide for using evaluation to improve informal education experiences. https://www.nisenet.org/catalog/team-based-inquiry-guide

Europlanet Evaluation Toolkit

Speaker: Anita Heward, Europlanet 2024 Research Infrastructure, University of Kent, UK

In this presentation, we will give a practical guide to the Europlanet Evaluation Toolkit (http://www.europlanet-eu.org/europlanet-evaluation-toolkit/), a resource that aims to empower outreach providers and educators in measuring and appraising the impact of their activities. The toolkit is intended to provide advice and resources that can be simply and easily integrated into normal outreach and education activities. The toolkit has been developed over a number of years with input from professional outreach evaluators and from active outreach providers within the planetary science community. The toolkit includes a brief introduction to evaluation, a choice of 14 data collection tools, worked examples of data analysis techniques, case studies and tutorials.

Talk link: https://youtu.be/d2bdL_oKMUA

Evaluation can provide essential information in understanding the effectiveness and accessibility of outreach activities in engaging diverse communities. In this presentation, we will give an overview of the Europlanet Evaluation Toolkit, a resource that aims to empower outreach providers and educators in measuring and appraising the impact of their activities. The toolkit is intended to provide advice and resources that can be simply and easily integrated into normal outreach and education activities. It is available as an interactive online resource at https://www.europlanet-society.org/outreach/europlanet-evaluation-toolkit/, as a downloadable PDF and as a hard copy (including a book and set of activity cards).

The toolkit has been developed over a number of years with content provided by professional outreach evaluators Karen Bultitude and Jennifer DeWitt (UCL, UK). Initially, a series of focus groups and scoping discussions were held with active outreach providers from the planetary science community, in order to determine what they wanted from such a toolkit, and what sort of tools would be of most interest. A shortlist of tools was developed based on these discussions, with volunteers testing out the tool instructions once they were drafted.

The toolkit begins with a brief introduction to evaluation and steps to choosing the right tools. This advice takes the form of a series of questions to help design an evaluation approach and make the most efficient and effective use possible of limited time and resources. The toolkit offers a choice of 14 data collection tools that can be selected according to the audience (e.g. primary, secondary, interested adult, general public), the type of environment and activity (e.g. drop-in, interactive workshop, ongoing series, lecture/presentation or online) or according to when they might best be used (during, beginning/end, or after an event). The online version of the toolkit includes a set of interactive tables to help with the selection of which tool is most appropriate for any given situation.

The data collection tools are:

- Physical Ranking Scales: https://www.europlanet-society.org/evaluation-tool-physical-ranking-scales/
- Graffiti Wall: https://www.europlanet-society.org/evaluation-tool-graffiti-wall/
- Mentimeter: https://www.europlanet-society.org/evaluation-tool-mentimeter/
- Palm on Chest: https://www.europlanet-society.org/evaluation-tool-open-palm-on-chest/
- Geographic Location Map: https://www.europlanet-society.org/outreach/europlanet-evaluation-toolkit/evaluation-tool-geographic-location-maps/
- Snapshot Interviews: https://www.europlanet-society.org/outreach/europlanet-evaluation-toolkit/evaluation-tool-snapshot-interviews/
- Pre/Post Quizzes: https://www.europlanet-society.org/outreach/europlanet-evaluation-toolkit/evaluation-tool-pre-post-quizzes/
- Pebbles in a Jar or Box: https://www.europlanet-society.org/outreach/europlanet-evaluation-toolkit/evaluation-tool-pebbles-in-a-jar/
- Three Words: https://www.europlanet-society.org/outreach/europlanet-evaluation-toolkit/evaluation-tool-three-words/
- Target Evaluation: https://www.europlanet-society.org/evaluation-tool-target-evaluation/
- Post Event Surveys: https://www.europlanet-society.org/evaluation-tool-post-event-surveys/
- Photograph Diary: https://www.europlanet-society.org/evaluation-tool-photograph-diary/
- Peer Interviews: https://www.europlanet-society.org/outreach/europlanet-evaluation-toolkit/evaluation-tool-peer-interviews/
- Tweet Sentiment Visualisation: https://www.europlanet-society.org/evaluation-tool-tweet-sentiment-visualisation/

The toolkit also includes descriptions and examples of how to use two techniques (word-clouds and thematic coding) to analyse the data, as well as some top tips for evaluation and recommended resources. For some of the tools, the case study examples include information about how the tools have been used in the context of an event, how the data was collected and

analysed, and what conclusions were reached, based on the data gathered. Over the past year, videos and training resources for using the toolkit have been added, as well as virtual alternatives to the physical tools. Case studies contributed by the community are very much welcome.

The Europlanet Evaluation Toolkit has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149 (Europlanet 2024 RI) and 654208 (Europlanet 2020 RI).

Assessing Students in the 21st Century

Speaker: Priscila Doran, NUCLIO / Project Coordinator, Portugal

Innovation in education has been a very much debated theme around the world and teachers have been making huge efforts to update their teaching methods. From shifting the classroom into a more student-centered environment to the use of games for learning and the personalized and individualized approach to teaching, education is suffering a huge transformation. However, one question is often asked by teachers, which is: "How can I assess my students when I teach in innovative ways?". When teachers shift their teaching methods and offer students a diverse and more personalized learning experience, the traditional standardized tests and exams become obsolete. If students encounter different learning opportunities inside the classroom and have the freedom to explore their own interests while learning fundamental life skills, it is natural to think that each student will acquire slightly different pieces of knowledge and retain different concepts at the same time. Therefore, a standardized assessment that focuses on such knowledge retention will not effectively portrait the real learning development of each student. Moreover, in an era where knowledge is easily accessed through a browser and a smartphone, it becomes imperative to shift the focus of student assessment. More than evaluating the ability to retain knowledge, it becomes important to focus both teaching and assessment in the development of fundamental skills like learning how to learn, critically thinking, innovation, divergent thinking, collaboration, communication, creativity, self-confidence, self-awareness, self-regulated learning, amongst others. Considering all the aforementioned it is important to develop innovative ways of assessing students so that assessment itself becomes a powerful tool in the learning process. Tools like checklists, rubrics and automated global assessment tools are proposed.

Talk link: https://youtu.be/aGE-8Ob3zmk

The system that we call today "traditional education" was once created to teach very specific skills and knowledge that would be required in very well-known lines of work where individuals would probably work throughout their lives [1]. However, the world has evolved and the very fast and accelerating technological development has changed the future perspectives of every student entering school, increasing the possibilities and unpredictability of their future.

Students entering school now will discover possibilities in their future that we might not imagine or predict today [2]. As a consequence, it becomes a futile attempt to try to determine what specific knowledge they need to acquire for their future career. Moreover, since knowledge today is readily available through one-click on a smartphone, rather than requiring students to memorize a whole list of concepts, it becomes much more important to focus on important core concepts and to teach students relevant skills for their future. Students in this century must learn how to learn, how to distinguish valid from non-valid information, how to think critically, how to be innovative and creative, how to communicate, how to be tolerant and respectful, etc. [3]

In order to achieve this, teachers have been making efforts worldwide to innovate in their teaching practice, joining innovative projects, using new tools, new methodologies and changing the classroom environment [4]. More importantly, education is becoming more inclusive and more personalized and this often raises an important question: "If we are offering a personalized learning experience, how can we assess our students"?.

When teachers shift their method into a student-centered approach and allow each student to explore learning at their own pace and according to their own interests and talents, it is natural that each student will learn differently and sometimes grasp different concepts. Furthermore, students will develop different skills and at a different pace. As such, using standardized assessment methods like tests and exams, that focus on memorized knowledge, becomes ineffective. Consequently, in order to properly innovate in their teaching, teachers also need to innovate in the way they assess their students.

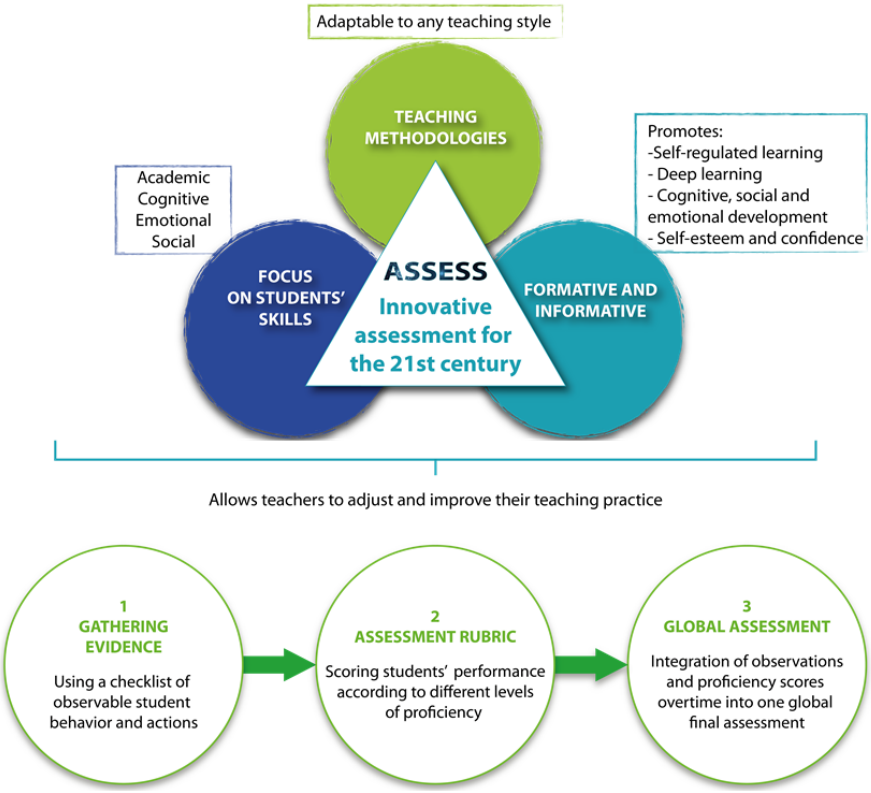
When used correctly, assessment can be a powerful tool for learning and for self-development, but in order to innovate and create a meaningful assessment for the 21st century, we first need to reflect on a few questions. The first question we should think about is: "why do we assess our students?" Do we assess them to give a final score, make them compete with each other and place them in a ranking that rewards them for their ability to memorize knowledge? Or should we assess students to provide them with the best support and self-awareness for the development of fundamental skills and invest in collaboration and acceptance instead of competition?

The second question is: "Are we being inclusive when we use a standardized assessment, like tests and exams?"

"Sometimes, the most brilliant minds do not shine in standardized tests, because they do not have standardized minds" - Diane Ravicht

Human beings are all different from each other and there is no such thing as a standard person. If we acknowledge this, then it becomes clear that assessment should not be standardized, but diverse, fluid and adaptable to each individual.

Student Assessment in the 21st Century



So, in order to solve the problem, we need to revolutionize assessment. And to do so, it is important to look at students' assessment from different perspectives. It is important to innovate not only the methods and tools that we use, but also the overall assessment, teaching and learning mindset.

We need to invest in an assessment methodology that is adaptable to any teaching style, that focuses on students' skills and that it is formative and informative, instead of judgmental. An assessment that is continuous throughout the learning process and that provides students with multiple opportunities for development and improvement. And through this, it also provides teachers the opportunity to continuously evaluate the effectiveness of their methods, as well as improve their teaching and adjust it to the needs of each class.

We need new tools that allow for an effortless and quick formative evaluation. In this presentation we propose tools like checklists, rubrics and an automatic global assessment that allow teachers and students to collect valuable information and have a visual representation of the development along the year. The checklists allow both students and teachers to regularly assess what has been achieved and what needs to be further developed. The rubric allows for a regular quantitative and qualitative evaluation of the proficiency and development of students in the different fields of assessment. And finally, the global assessment tool gathers all the information in one place and automatically creates graphics of each student's development in each field of assessment.

Finally, we need a completely new mindset on assessment and in the school as whole, where assessment is considered a powerful tool for the development of fundamental life skills. A mindset that promotes collaboration, tolerance and acceptance and where each student is treated and accepted as a unique human being and is valued by their qualities and talents.

NUCLIO is coordinating a project (The ASSESS project: https://assess.nuclio.org) that focuses entirely on the creation of an innovative mindset around student assessment and that will integrate innovative assessment tools into a digital app for teachers and students. We welcome all teachers, educators and all those who are interested to participate in the project, to share your ideas with us and to give us your contribution.

To have access and explore some new assessment tools, teachers can explore the assessment toolkits that were designed in the framework of two Erasmus+ projects: the IDiverSE assessment toolkit (https://idiverse.eu/idiverse-assessment-toolkit) and the POLAR STAR assessment toolkit (http://polar-star.ea.gr/content/assessment-toolkit).

References:

1. Moving Our Education System Forward (2020). Factory Model of Education. Available online at: http://creativecurriculumisabella.weebly.com/the-factory-model-of-education.html (accessed September 20, 2021)

2. European Commission (2017) 'White Paper on the future of Europe': https://europa.eu/european-union/sites/europaeu/files/whitepaper_en.pdf

3. Howells, K. (2018). The future of education and skills: education 2030: the future we want, OECD: https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf

4. Stéphan, V. L., Joaquin, U., Soumyajit, K., & Gwénaël, J. (2019). Educational Research and Innovation Measuring Innovation in Education 2019 What Has Changed in the Classroom?: What Has Changed in the Classroom?. OECD Publishing.

How did that happen? Mixed-method Evaluation of Astronomy Resources

Speaker: Sophie Bartlett, Cardiff University / Faulkes Telescope Project, UK

Astronomy is a well-known effective point of engagement, sparking students' curiosity and desire to learn. However, astronomy has a lot to offer; pretty pictures, a mind-blowing vastness, big telescopes, and application of much of the more mundane classroom science. As a result, it can be difficult to disentangle what specifically causes students' engagement. That is, what does an astronomy resource need in order to be effective and what other components are surplus to requirement? This presentation focuses on a PhD study involving 226 secondary school students that set out to answer this question. Focusing on both methodology and results, this presentation will explore how mixed-method evaluation can offer valuable information for developing and delivering effective astronomy resources. By using quantitative methods to identify what happens, and qualitative methods to identify why this happens and under what circumstances. Although mixed methods evaluation demands greater time and manpower, it can provide hugely valuable results that are not exclusive to a single resource, but that provide transferable findings that can be of use to future development and wider educationalists.

Talk link: https://youtu.be/VK_XnMB6wec

– –

Astronomy is a well-known effective point of engagement, sparking students' curiosity and desire to learn (Salimpour et al., 2021; Osborne and Collins, 2001). However, astronomy has a lot to offer; pretty pictures, a mind-blowing vastness, big telescopes, and application of much of the more mundane classroom science. As a result, it can be difficult to disentangle what specifically causes students' engagement and positive learning experience when engaging with such materials. That is, what does an astronomy resource need in order to be effective and what other components are surplus to requirement?

A mixed method approach to evaluating such resources offers a valuable opportunity to capture this information. Its opportunities and strengths are argued here in the context of a PhD study involving ten case studies of secondary school classrooms that set out to identify how teacher-implemented astronomy resources can promote student learning experiences.

In a broad sense, evaluation is typically quantitative or qualitative. Quantitative approaches are used to explain a particular phenomenon through numerical data collection. They are deductive, objective and outcome-oriented. Qualitative approaches are inductive, subjective and process-oriented (Streefkerk, 2021). Each approach has its own advantages and disadvantages (see Streefkerk, 2021 for further detail). However, by combining the two methods we are able to surmount many of the individual limitations and utilise the individual advantages. As a result,

the evaluator can gain a more holistic understanding of the phenomena at study. Where the quantitative data can provide an understanding of *what* happened, the qualitative data can give a more comprehensive understanding of *why* it happened (Johnson and Onwuegbuzie, 2004; Ivankova et al., 2006; Denscombe, 2008; Greene, 2008). This is particularly useful in formative evaluation, where you are looking to develop or improve resources or where you yielded negative responses. Evidence of poor or even negative impact in quantitative data is disheartening, but not understanding why such a result was yielded due to a methodological limitation can also leave you with little understanding of what caused such a result and what steps can be taken to improve results in the future. Although mixed methods evaluation demands greater time and manpower, it can provide hugely valuable results that are not exclusive to a single resource, but that provide transferable findings that can be of use to future development and wider educationalists.

Creswell and Plano-Clark (2011) also discuss the sequencing of mixed methods. They explain that depending on what method you implement first, your evaluation can be either explanatory or exploratory. Exploratory methods begin with qualitative methods and then generalisations are sought through a follow-up quantitative method. This is helpful if you are looking to make generalisations, perhaps if your resource was effective among a small student cohort, you would want to identify if it is effective among larger or additional cohorts.

Alternatively, explanatory methods first apply a quantitative phase in order to understand the general picture, and is followed by a qualitative phase in order to explore the patterns from the quantitative data and why such findings were yielded. An explanatory approach is helpful when you want to understand the processes and mechanisms behind the quantitative results, perhaps to inform future astronomy resource development.

In the case of this PhD study, a sequential explanatory design was followed as the researcher wanted to understand specifically what happened but also what processes and experiences led to those outcomes so that findings could be transferred to future and wider educational resource development. The quantitative method involved a closed-questionnaire consisting of Likert-scale items. The questionnaire was implemented on two occasions: before students had engaged with the astronomy resources, to reflect on these five areas in relation to their day-to-day science lessons, and after students had engaged with the resources, to reflect on their experiences when using the astronomy resources. The use of a parallel questionnaire pre- and post-engagement meant that direct comparisons could be drawn and also offered a tool that could be used with future resources.

For the qualitative arm, classroom observations, student focus groups and teacher interviews were implemented with a smaller subset of the audience. Classroom observations were carried out before astronomy resource implementation (during a 'normal' science lesson) and while students were using the astronomy resources. Focus groups and interviews were implemented after implementation of the activities. In line with the sequential explanatory design, the focus group and interview question schedules were informed and guided by preliminary results from the questionnaires and observations. This allowed the researcher to explore why such events took place and why students were or were not engaged.

This process of mixed method evaluation revealed five key elements that promoted positive learning experiences among students. Although these were identified in the context of particular

astronomy resources, these elements are considered applicable to any activity or resource. Each of the five elements is now described, with accompanying relevant quote excerpts from students. 1. Processes of **investigation and exploration** that encourage students to follow the scientific process of gaining new knowledge: *"It was interactive, using real data. This makes it feel more relevant"*. 2. **Experiences of autonomy** gives students a sense of ownership over their learning: *"I liked the freedom of finding things out on our own"*. 3. **Novel, unexpected experiences** offer a 'wow' factor and provide an element of surprise: *"What? So that is the age of the Universe? I feel like Einstein!"*. 4. Providing students with opportunities to **cooperate and collaborate** with their peers helps provide a sense of relatedness (Gagne and Deci, 2005) *"It is easier to work off each other, like some people might have stronger points in that subject so they can teach other people stuff"*. 5. Embedding **effective differentiation** into resources to foster students' confidence and provide them with a challenging but achievable task: *"So I can do science"*.

However, despite the opportunities for autonomy and a student-centred classroom, the role of the teacher was still crucial in influencing students' experiences. Observations of each classroom revealed that despite using the same resource, implementation differed in each setting. Where teachers had a great awareness of individual learning needs among students, they were able to adapt the resources and embed appropriate differentiation. Additionally, great preparation from the teacher and familiarity with the resource led to more positive learning experiences among the students.

The results of this study provided valuable insight for educators and resource developers. The five key elements that were seen to promote positive learning experiences can stand as a foundation when developing a resource. Resource developers should recognise that resources will be implemented slightly differently in every classroom and thus should consider the teacher's role and their support needs. Resources should be adaptable and apply various scaffolds that can be added or removed in order to differentiate appropriately.

References:

- Creswell, J. and Plano-Clark, V. 2011. Designing and Conducting Mixed Methods Research. 2nd Ed. Los Angeles, US: SAGE Publications

- Denscombe, M. 2008. Communities of Practice: A Research Paradigm for the Mixed Methods Approach, Journal of Mixed Methods Research, 2(3), pp. 270-283. DOI: 10.1177/1558689808316807

- Gagne, M. and Deci, E. 2005. Self-Determination Theory and Work Motivation, Journal of Organizational Behaviour, 26, pp. 331-362. DOI: 10.1002/job.322

- Greene, J. 2008. Is Mixed Methods Social Inquiry a Distinctive Methodology?. Journal of Mixed Methods Research, 2(1), pp. 7-22. DOI: 10.1177/1558689807309969

- Ivankova, N., Creswell, J. and Stick, S. 2006. Using mixed-methods sequential explanatory design: From theory to practice, Field Methods, 18(1), pp. 3-20. DOI: 10.1177/1525822X05282260

- Johnson, R. and Onwuegbuzie, A. 2004. Mixed Methods Research: A Research Paradigm

Whose Time Has Come, Educational Researcher, 33(7), pp. 14-26.
DOI: 10.3102/0013189X033007014

- Osborne, J. and Collins, S. 2001. Pupils' views of the role and value of the science curriculum: A focus-group study, International Journal of Science Education, 23(5), pp. 441-67.
DOI: 10.1080/09500690010006518

- Salimpour, S et al. 2021. The gateway science: a review of astronomy in the OECD school curricula, including China and South Africa. Research in Science Education 51, pp. 975-996.
DOI: 10.1007/s11165-020-09922-0

- Streefkerk, R. 2021. Qualitative vs. Quantitative Research. Available at: https://www.scribbr.com/methodology/qualitative-quantitative-research/#:~:text=and%20qualitative%20methods%3F-,Quantitative%20research%20deals%20with%20numbers%20and%20statistics%2C%20while%20qualitative%20research,and%20experiences%20in%20more%20detail. [Accessed: 08 November 2021]

Discussion summary

The presence in both sessions of all the speakers created the basis of very fruitful discussions about different aspects of evaluation. They stressed the importance of planning evaluation from the very beginning in the project design (even for short projects), in order to make the experience as positive, and its results more useful, as possible.

We discussed observation protocols in the evaluation process, in particular for young children, besides interviews and drawing pictures, pointing out the importance of evaluating the students' engagement level.

Similarly, it is important to note that when designing surveys with many types of questions, it is sometimes difficult to balance between a reasonable survey length and reliable results. Indeed, if the survey is too long, and it presents many open-ended questions, it could discourage people from compiling it. A suggestion is to verify that the duration of the compilation does not take more than about ten minutes.

Finally, we discussed the difference between evaluation and research: even if they are similar in data collection and analysis, they are different in their overall purpose. Research is about understanding phenomena and generalizing, while evaluation is about drawing judgement conclusions about quality, merit or worth. For both, anyway, it is important to respect some ethics requirements in data gathering and using. Some ethical guidelines have been developed by the American Evaluation Association (see https://www.eval.org/About/Guiding-Principles).

http://astro4edu.org

@astro4edu

IAU OFFICE OF ASTRONOMY FOR EDUCATION

OAE

THE SHAW PRIZE 邵逸夫獎

IAU