

Monitoring and evaluating astronomy outreach programmes: Challenges and solutions

AUTHORS:

Sarah Chapman¹
Laure Catala^{2,3,4}
Jean-Christophe Mauduit⁴
Kevin Govender⁴
Joha Louw-Potgieter¹

AFFILIATION:

¹Institute for Monitoring and Evaluation, Section for Organisational Psychology, The School of Management Studies, University of Cape Town, Cape Town, South Africa

²Department of Astronomy, University of Cape Town, Cape Town, South Africa

³South African Astronomical Observatory, Cape Town, South Africa

⁴International Astronomical Union, Office of Astronomy for Development, Cape Town, South Africa

CORRESPONDENCE TO:

Joha Louw-Potgieter

EMAIL:

Joha.Louw-Potgieter@uct.ac.za

POSTAL ADDRESS:

Institute for Monitoring and Evaluation, Section for Organisational Psychology, The School of Management Studies, University of Cape Town, Office 4.40 Leslie Commerce, Upper Campus, Rondebosch 7708, South Africa

DATES:

Received: 02 Apr. 2014

Revised: 15 Aug. 2014

Accepted: 04 Sep. 2014

KEYWORDS:

astronomy for development; STEM education; science outreach; programme evaluation; monitoring and evaluation framework

HOW TO CITE:

Chapman S, Catala L, Mauduit J-C, Govender K, Louw-Potgieter J. Monitoring and evaluating astronomy outreach programmes: Challenges and solutions. *S Afr J Sci.* 2015;111(5/6), Art. #2014-0112, 9 pages. <http://dx.doi.org/10.17159/sajs.2015/20140112>

A number of tools exist to guide the monitoring and evaluation of science, technology, engineering and mathematics (STEM) education and outreach programmes. Fewer tools exist for evaluating astronomy outreach programmes. In this paper we try to overcome this limitation by presenting a monitoring and evaluation framework developed for the International Astronomical Union's Office of Astronomy for Development (OAD). The mandate of the OAD is to stimulate sustainable development at an international level and to expand astronomy education and outreach globally. The broad assumptions of this programme are that astronomy has the potential to contribute to human development by means of the transferable nature of its science discoveries, as well as its potential to activate feelings of wonderment, inspiration and awareness of the universe. As a result, the programme potentially embodies a far broader mix of outcomes than conventionally considered in STEM evaluation approaches. Towards this aim, we operationalise our monitoring and evaluation approach by first outlining programme theories for three key OAD programmes: a programme for universities and research, another one for schools, and one for public outreach. We then identify outcomes, indicators and measures for each one of these programmes. We conclude with suggestions for evaluating the global impact of astronomy for development.

Introduction

What does gazing at the stars and putting a man on the moon have to do with monitoring and evaluation?

The answer lies in a spate of recent discussions around evaluation of science, technology, engineering and mathematics (STEM) interventions. At the 27th Annual Conference of the American Evaluation Association in 2013, the STEM Topical Interest Group emerged as a fully fledged area of interest and sponsored 23 different sessions. Apart from presentations dealing with evaluations of STEM education, delegates affiliated to the National Aeronautics and Space Administration (NASA) presented a paper entitled 'Measuring Inspiration' to describe their agency-wide approach to advance high-quality STEM education.¹ In their presentation, they highlighted the need for rigorous and thorough performance assessment of astronomy outreach programmes, which they defined as evaluation approaches that include both an assessment of how well the programmes are being implemented (process evaluation) and whether they are achieving their aims (outcome assessment). Ideally, these assessments should be based on plausible theories on how the programme is supposed to work (i.e. a programme theory), and use reliable and valid data collection instruments.¹

In 2011, the International Astronomical Union's (IAU) Office of Astronomy for Development (OAD) was opened at the South African Astronomical Observatory in Cape Town with the mandate to stimulate development at an international level and expand astronomy education and outreach globally. The formative and emergent nature of the programme meant that simple, focused and very practical monitoring and evaluation approaches were needed. In this article, we describe the method we used to develop a programme monitoring and evaluation (M&E) framework for the OAD.

Astronomy for development

The IAU is an international astronomical organisation of more than 10 000 professional astronomers from more than 90 countries. Its mission is to promote and safeguard the science of astronomy in all its aspects through international cooperation. The International Year of Astronomy (2009) inspired the IAU to 'commit to even more ambitious programmes of educating the world to the beauty of the Universe and the sense of common humanity that derives from it'^{2(p.3)}. The IAU's vision, through the establishment of the OAD, was to promote human development by means of astronomy outreach. Outreach activities were focused around three core programme areas: (1) universities and research, (2) children and schools and (3) the public.

The OAD's vision that astronomy has the potential to lead to positive human development was underscored by a number of hypotheses or core assumptions. The first of these was that because astronomy has triggered curiosity and a certain form of fascination throughout cultures, continents and generations, it has the ability to reach out to as broad an audience as possible. This reach, in turn, enables astronomy to plant the seeds for a sense of common heritage and a shared, overarching (or superordinate) sense of humanity – building blocks for a generally more tolerant society.

A second hypothesis was that astronomy has the potential to contribute to human development by means of the transferable nature of its developments and discoveries. In support of this assumption, the OAD cites a number of technologies and skills developed for astronomical research that are now applied in industry, the medical field, and in devices that one uses on a daily basis.³ A few examples of these are the charge-coupled devices (better known as CCDs) that one finds in digital cameras and cellular phones; the transfer of technology developed for astronomy

© 2015. The Author(s).

Published under a Creative Commons Attribution Licence.

into medical imaging instrumentation; and the application of adaptive optics technology in the high-precision laser industry and medicine.

A final assumption was that astronomy as a science has the potential to positively build the economic, institutional and human capital of participating countries and institutions. For example, according to Schilizzi et al.⁴, the square kilometre array radio telescope's influence will be widely felt in astroparticle physics and cosmology, fundamental physics, galactic and extragalactic astronomy, solar system science and astrobiology. A report from Promoting Africa European Research Infrastructure Partnerships (PAERIP)⁵ – a European Union–Africa partnership – elaborates on the positive socio-economic impact of a large astronomical project such as the Southern African Large Telescope (SALT). Via the SALT collateral benefits programme, the building of SALT is believed to have generated employment, skills and human-capacity building, science awareness in the surrounding communities, as well as the development of teacher-training and higher-education programmes. More broadly, a recent article in *Nature*⁶ highlights the importance of supporting science and ways of using it for capacity building.

Evaluating astronomy for development

A key question for the OAD was the extent to which their assumptions about impact (outlined above) were plausible, and could be substantiated by empirical evidence stemming from their outreach activities. For this reason, OAD early on identified the need for their outreach activities to be underscored by a robust impact evaluation platform.

But evaluators are not able to determine impact with certainty, only with varying degrees of confidence.⁷ Degrees of confidence in turn are dependent to a large extent on the research design utilised for estimating programme effects, how distal the outcomes of interest are, and how mature the programme of interest is. The OAD is not a mature programme, and the outcome of interest – human development – is both distal and difficult to define. If evaluators could, for example, compare participation in scientific research within countries where astronomy for development programmes were implemented with participation in countries where no programmes were implemented, one could theoretically estimate the effect difference with relative certainty. However, implementing such a design might call for randomised field experiments or the selection of carefully matched comparison sites to enable a quasi-experimental design. This scenario presents challenges in terms of time, cost, accurate data and cooperation.

A less robust approach is the measure of 'perceptions' of impact. This measure consists of gathering anecdotal testimonies of how a project is perceived to have affected the lives of its target audience, and thus contributed to the intended outcomes. The data can be in the form of narratives, pictures or short films. Although this approach may be satisfying to some audiences, the true measure of impact is likely to remain as an elusive end goal.

In light of these difficulties, we made a strategic decision when defining our M&E approach to not focus on measuring distal impacts; that is, testing the OAD's pre-suppositions relating to human development outcomes. Rather, the decision was made to place initial emphasis on ensuring effective implementation and short-term outcome attainment for public, educational and university outreach programmes. The rationale behind this decision was that if the OAD's outreach activities were not being implemented effectively or achieving even their short-term aims, more distal impact pathways – whether plausible or not – are unlikely to ever be realised. In our paper, we therefore restrict our M&E approach to an assessment of the design, implementation and short-term outcomes of the OAD's astronomy outreach programmes.

For our purposes, we understand programme evaluation to be defined as 'the use of social science research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action in ways to improve social conditions'^{7(p.431)}, and programme monitoring as 'the systematic documentation of aspects of program performance that are indicative as to whether the program is functioning as intended'^{7(p.171)}. An M&E approach is thus an approach

that uses social science research methods to systematically document programme performance and functioning with a view to inferring the extent to which the programme is improving the social conditions of target beneficiaries. Typically, programme monitoring will focus on the continuous measurement of programme implementation and outcomes, whereas evaluation efforts can be more broad-based in that they might also assess, for example, the design of the programme, the need for the programme, and the cost-effectiveness or efficiency of the programme.

Challenges in evaluating astronomy outreach programmes

In developed countries, aspects of astronomy are frequently integrated into formal school curricula in an effort to inspire and interest students in science⁸, and numerous studies exist in which researchers have examined how learners' conceptual models of astronomical phenomena might best be moved towards more scientific notions⁹⁻¹⁰. In contrast, the benefits of integrating astronomy into informal public, educational and university outreach programmes are less clear. Although there have been calls for M&E approaches that specifically enhance the design and delivery of astronomy outreach,¹¹ most of these programmes have been viewed in terms of the degree to which they contribute to STEM development, and have been evaluated accordingly.

Educational STEM programmes are relatively straightforward to evaluate and a number of resources are available in the literature.¹²⁻¹⁴ Short- and medium-term outcomes typically are defined in terms of so-called 'STEM activation', which usually includes curiosity towards STEM, awareness of scientific principles, self-efficacy in STEM, and belief in the importance of the scientific enterprise. Long-term outcomes usually relate to competence, knowledge and mastery of STEM-relevant skills.¹² For tertiary-level STEM programmes, increased collaboration with university scientists may be an important additional outcome.¹⁵

While astronomy outreach programmes share many objectives with STEM programmes, there are a number of other, less understood areas through which we hypothesised that astronomy programmes may hold unique benefits. In astronomy we are dealing with a science that explores the universe and celestial objects within it. For this reason, it is an inspirational science that has the potential to evoke feelings of wonderment and a yearning to understand our origins. Our fascination with the universe may even result in a sense of oneness that contradicts and undermines those national and cultural boundaries that separate us. For these reasons, we felt that astronomy outreach potentially embodies a far broader mix of outcomes than those conventionally considered in most STEM evaluation approaches.¹² And while some evaluation frameworks for assessing learning outcomes (such as the Generic Learning Outcomes framework¹⁶) acknowledge constructs such as enjoyment, creativity and inspiration in their conceptual frameworks, in order to operationalise our M&E framework, we were faced with the challenge of providing indicators and measures for such outcomes.

Approaches to astronomy outreach evaluation have for the most part failed to reflect this complexity. In one example from Hawaii, the Imoloa Astronomy Center in Mauna Kea claimed a role in the mitigation of cultural differences between astronomers and the Hawaiian community. This role, however, was never formally assessed by means of an empirical M&E approach.¹⁷ The well-known European Universe Awareness programme also makes it clear that project goals include the somewhat abstract outcome of changes in intergroup attitudes, but no guidelines as to how these domains might be measured are provided in their evaluation manual.¹⁸ Rather, evaluation tends to be directed solely at the level of astronomy awareness, knowledge, understanding and skills,^{19,20} for example, correctly identifying a galaxy (knowledge) or using a telescope (skills). A final example is provided in the evaluation approach of the Sol programme run by NASA for underrepresented communities.²¹ In the evaluation of the pilot of this programme, outcomes were expressed solely in terms of students' positive opinions and knowledge of STEM fields and careers (not specifically astronomy), as well as performance in science and mathematics classes. For this evaluation, generic survey tools relating to general STEM activation were adapted by the M&E team.

Although guidelines on evaluating attitude and behaviour change related to informal science programmes are available in the literature,²² they are fairly broad and will need to be adapted and developed in more depth for specific astronomy-related projects.

Empirical evidence for the effects of astronomy outreach programmes

Overall, there is a lack of empirical evidence as to the effects of astronomy outreach programmes on human development. Although studies make it clear that astronomy frequently rates higher in the public mind than other science subjects on the basis of it encapsulating abstract concepts such as ‘remoteness’, ‘unknownness’, and ‘excitement of discovery’^{123(p.225)}, whether these sentiments can be useful to human development is less definite. Indeed, it is not even clear if educational astronomy programmes are even effective in inspiring lasting interest in astronomy, let alone broader areas of science and development.

One study examined the experience of 655 10- and 11-year-olds in the United Kingdom who took part in a simulated space trip.²⁴ Although a quarter of the children were inspired by the visit to become scientists, half showed no significant changes and some even showed negative changes in their attitude. A later study²⁵ used a pre-test, multiple post-test design to assess the lasting effects of a space centre visit on elementary school children’s attitudes towards science and astronomy. Although over 90% of students who visited the centre were highly excited by astronomy after completing the visit (a quarter expressed a desire to become astronauts one day), there was no evidence of the visit having a statistically significant effect on children’s enthusiasm for science. Improvements in children’s views about science and being scientists in future were also marginal, and over a period of a few months declined to the point that the final post-test means were only slightly higher than the pre-visit means.

Studies of astronomy training at a tertiary level have shown similar trends. One evaluation tracked three cohorts of over 400 Mexican students enrolled in a semester-long introductory astronomy course.²⁶ Although students typically progressed well in terms of conceptual subject mastery, there was little to no significant change over each semester in students’ positive attitudes about astronomy specifically, and science generally. Teachers’ attitudes may be even harder to influence than students. Ucar and Demircioglu²⁷, for example, reported that a semester-long astronomy course did not change teacher attitudes toward astronomy, and that only marginal gains were evident after even a full 4-year programme.

Given these challenges, the value added by exposing learners, students and the public to astronomy (as opposed to STEM outreach more generally) needs to be carefully considered.²⁵ Research has suggested

that, like any extracurricular science outreach programme, informal astronomy outreach programmes – especially those targeted at children – require skilful facilitation and careful integration of content into school curricula if attitudes and learning are to be positively influenced.^{25,28,29} And while more formal, tertiary-level programmes may be effective in facilitating conceptual mastery of astronomy principles,³⁰ considerable work needs to be done if students are to be inspired to pursue STEM subjects generally (and astronomy career paths specifically) as a result of completing these courses. For programmes such as the OAD’s, where a lack of facilities and resources in developing countries might understandably limit programme quality, these challenges should be taken particularly seriously. There is therefore a clear need for simple yet effective formative M&E systems that are properly aligned to the programme’s impact theory.

A programme theory for the astronomy outreach programmes

We used a theory-based approach to develop our M&E framework.^{7,31-36} The approach involves the evaluators interacting with the programme stakeholders to draw out their programme theory until the stakeholders ‘find little to criticize in the description’⁷. M&E can then be focused on ensuring that benchmark processes and outcomes in the programme’s stated theory are being met.

A programme theory can be simply defined as ‘a plausible and sensible model of how a programme is supposed to work’^{34(p.5)}, or more specifically as ‘the set of assumptions about the manner in which a program is related to the social benefits it is expected to produce and the strategy and tactics the program has adopted to achieve its goals and impacts’^{7(p.432)}. Rossi et al.⁷ go on to define two components of programme theory – the programme’s process theory, which outlines the assumptions and expectations about how the programme is expected to operate, and the impact theory, which describes the cause-and-effect sequences brought about by programme activities which lead to programme impact. This distinction is similar to the distinction between action and change theory made by Chen³³, who describes programme theory as a set of stakeholders’ prescriptive assumptions on what actions are required to solve a problem (i.e. an action theory), as well as descriptive assumptions about why the problem will respond to the actions (i.e. a change theory).

We started to build our M&E framework by developing a high-level model for the three programmes of the OAD (universities and research, children and schools, and public outreach). This model is depicted as a variable-oriented programme theory in Figure 1.

The assumption underlying this programme is: if programme activities for universities and research, children and schools, and the public are

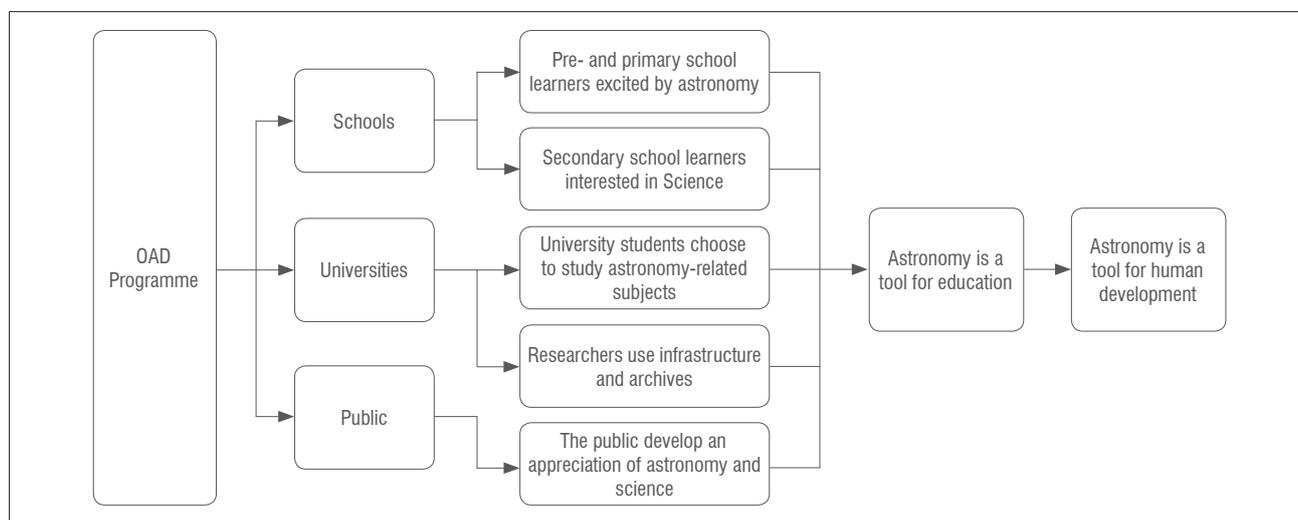


Figure 1: High-level variable-oriented programme theory for the Office of Astronomy for Development (OAD).

offered as intended and at the right intensity, astronomy will serve as a tool for education, and, in the long run, as a tool for human development.

Thereafter, we expanded the programme theories for each one of the three different programmes. We then added tables detailing the process and outcome indicators and measures to assess whether benchmark processes and outcomes are being achieved as planned. Based on these tables, we then developed data collection templates. This final step is specifically important for developing countries for which data collection and management may not be the norm. The templates also serve to ensure comparable data across projects, countries and continents. The OAD will require that programme managers use data from these templates when submitting regular progress reports. The templates will be available to programme managers in online or hard copy format, depending on technological standards within the relevant country. For the sake of brevity, the templates are not included here, but are available from the corresponding author and the OAD website.³⁵ However, the three surveys included appear in the data collection templates.

Programme theory for the Universities and Research Programme

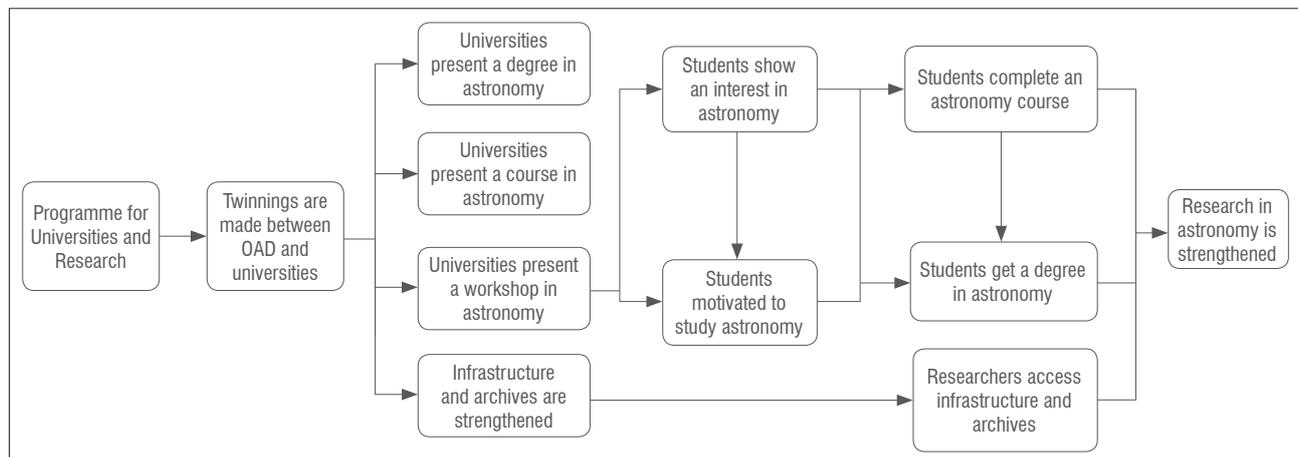
This programme can be considered a core programme of the OAD. Its goal is to create cohorts of graduates and researchers, in astronomy

specifically and in science generally. The programme theory is depicted in Figure 2.

Here the assumption is, that given the relevant offering at a university, students will become motivated to follow this field of study. As a result of this motivation, students will be more likely to study astronomy-related subjects and/or pursue a degree in astronomy.

The intervention involves twinning between universities, particularly in developing countries.² Typically, twinning involves mentorship of emerging researchers and students by senior international researchers, and the presentation of one or more workshops at a university that is not yet offering astronomy degrees or astronomy-related subjects. The programme also encourages researchers in astronomy and astronomy-related subjects to establish and access astronomy infrastructure and archives.

Data monitoring for this programme focuses on both process and outcome monitoring. Process monitoring primarily involves tracking the amount, type and quality of university twinning, as well as workshop presentations. For most of these indicators, programme records would typically serve as measures. Developing indicators and measures for the quality and strength of inter-institutional collaboration proved more challenging. We considered proxy indicators such as email correspondence (measure: correspondence frequency), but ultimately opted to use the presence of a memorandum of understanding as a measure for twinning amount and



OAD, Office of Astronomy for Development

Figure 2: Programme theory for the Programme for Universities and Research.

Table 1: Process and outcome indicators and measures for the Programme for Universities and Research

Process/Outcome	Indicators	Measures	Data source
Twinning made (process)	Amount, type, quality of twinning	Number and type of connections, signed Memorandum of Understanding, participant satisfaction, etc.	Organisational records, communication reports, participant feedback ratings
Workshops presented (process)	Amount, type, quality of workshops presented	Number and type of workshops, participant satisfaction, etc.	Participant feedback ratings
Universities present degrees/courses in astronomy (outcome)	Astronomy degrees and courses	Number of new courses/degrees etc.	Annual university survey
Students show an interest in astronomy and astronomy-related subjects (outcome)	Student attitude	Student ratings on items in workshop participant survey	Workshop participant survey
Students choose to study astronomy and astronomy-related subjects (outcome)	Astronomy degree and subject uptake	Students registered for target subjects/degrees	Annual university survey, workshop report
Researchers access astronomy archives and services (outcome)	Access to Astronomy Database Services	Number of people accessing, type of access	Annual university survey
Astronomy-related research increases (outcome)	Astronomy-related journal publications	Number, type of publication, etc.	Annual university survey

type. Participant satisfaction surveys were used to measure twinning quality. In contrast, outcome monitoring for this programme focused mainly on updating university records in order to track coverage in terms of astronomy and science graduates (Table 1).

We assumed that studying astronomy or science indicated an interest in these subjects and decided not to measure interest in astronomy for this programme. Where workshops in astronomy are presented at universities that do not offer courses or degrees in astronomy, the quality of the workshop presentation and the intention of workshop delegates to pursue a degree in astronomy or science is measured by means of a short survey. Items for this survey are presented in Table 2.

From Table 2, it is clear that many of the survey statements are quite blunt in nature, and do not include mechanisms for potentially increasing validity such as negatively worded statements. This bluntness was intentional. The global nature of the OAD programme meant that simple tools were needed that are unlikely to pose significant challenges to translation and analysis. Moreover, because respondents in some target countries are likely to be less familiar with surveys than their counterparts, the surveys needed to be both brief and direct in nature. Indeed, at this stage of the process of evaluating astronomy for development we are more concerned with whether these surveys are applicable to a specific target population across a wide range of socio-economic and educational contexts than with their content and construct validity. At a later stage, the psychometric properties of the surveys will need to be tested more rigorously.

Table 2: Participant survey to measure student attitude towards astronomy

1	I want to learn more about astronomy
2	I think I may become an astronomer one day
3	Astronomy is relevant to what I experience in my life
4	Astronomy is useful for making the world a better place
5	Attending this workshop has made it more likely that I will study astronomy at university
6	Attending this workshop has changed my perspective on how useful physics is as a subject
7	There is no astronomy at my university, but I would like to study physics

Response ranges from 1 (strongly disagree) to 4 (strongly agree), except for Item 7.

Programme theory for the Children and Schools Programme

The main goal of this programme is to activate excitement for astronomy in primary school children and to foster interest in science in secondary school children. However, several researchers have found that teachers often lack even the most basic scientific understanding of astronomical concepts,^{36,37} which suggests the need for improved training at this level. In light of these considerations, the OAD programme was disaggregated to two levels of programme beneficiaries: teachers (the direct recipients of OAD support) and learners (the secondary beneficiaries). The relevant programme theory is depicted in Figure 3.

The assumptions underlying this programme theory are: if relevant teachers are qualified (trained) to present and implement the programmes for schools as intended, the children who receive these programmes will be inspired by them; and older children will become interested in science.

In order to collect comparable data for process and outcome monitoring, we created a comprehensive list of indicators and suggested measures. A simplified version of this list is presented in Table 3. From Table 3, we can see that the outcome 'inspired by astronomy' is expressed at two levels of specificity: excitement in astronomy and identification with a superordinate grouping. Here, superordinate grouping identification is defined in terms of a sense of common humanity that transcends common sub-groupings such as national identity or ethnicity. These levels of specificity were developed with consideration to the OAD's strategic plan,⁸ as well as the personal experience of OAD staff during the first few years of project implementation. Excitement in astronomy might be measured by items relating to interest in astronomy and motivation to learn more about astronomy. In contrast, the outcome 'interest in science' is interpreted in terms of STEM activation. For our purposes, measures for STEM activation include items relating to interest in STEM, belief in the importance of STEM, and motivation to study STEM.

Two surveys are mentioned in Table 3: the primary and the secondary student surveys. When developing the primary school survey items, we had to keep in mind that in many developing countries primary school attendance is low and science often is not taught at this level. On the other hand, in many developed countries, children are taught science in sophisticated laboratories and start developing aspirations for careers as scientists at a young age. Because the OAD's mandate aims to establish impact within a global community that includes the developing world, measurable outcomes needed to be suitable for both developed and developing countries.

One of the most significant challenges was the lack of appropriate tools to guide the collection of M&E data in developing country contexts. While some STEM researchers have called for sensitivity to diversity, equality and cultural concerns in STEM programme evaluation^{38,39} the majority of STEM programmes do not consider these elements to be important

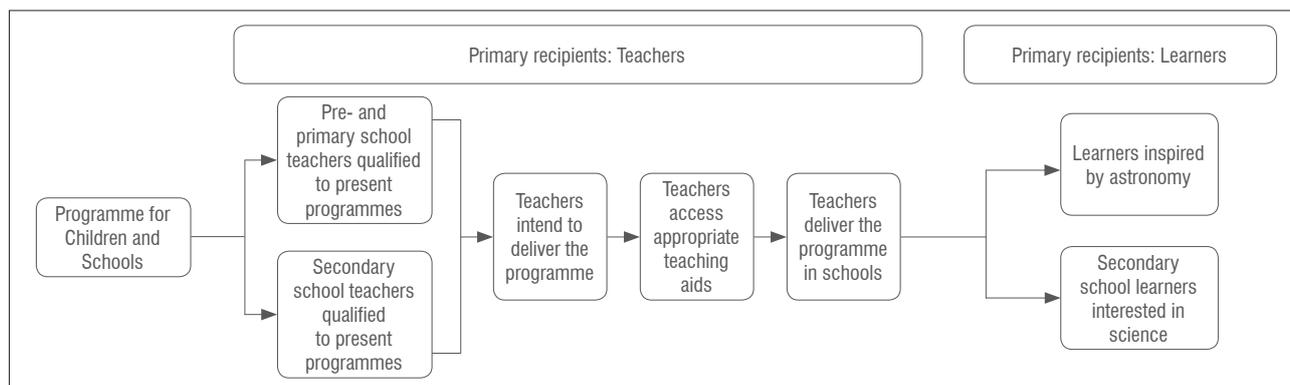


Figure 3: Programme theory for the Programme for Children and Schools.

to programme outcomes. Indeed, most researchers who work in this field have focused on race/ethnic disparity of minorities in developed countries,^{40,41} and there is marked absence of debate around adaption of STEM instruments to the developing world. In one study, for example, the effects of astronomy outreach programmes on children were measured by adapting existing scales that probe children's science enthusiasm as well as beliefs as to the value of science to society. Because science is not always offered (at least in practice) at the elementary level in some developing countries, measuring outcomes in terms of, for example, science enthusiasm and outlook on science,²⁵ seemed implausible. Other tools, such as the Survey of Attitudes towards Astronomy tool developed

by Zeilik et al.²⁶, have only been applied to university-level students.^{26,42} In contrast, tools more suitable for children of elementary school going age, such as the Space Interest Scale referred to by Jarvis and Pell²⁵, seemed too narrowly targeted at specific types of astronomy outreach (in this case a visit to a high technology simulation space centre).

Similar problems were encountered when developing the secondary school survey. Although we came across well-constructed and comprehensive surveys aimed at probing changes in attitudes towards STEM,⁴³ only a few of the items in these surveys seemed suitable for developing countries. Other tools we reviewed were problematic in

Table 3: Process and outcome indicators and measures for the Programme for Children and Schools

Process/Outcome	Indicators	Measures	Data sources
Teachers qualified (process)	Teacher training outputs	Number trained, type trained, nationality, etc.	Teacher training report
	Teacher training quality	Average facilitator ratings, certification, teacher assessment, etc.	Teacher survey
Teachers equipped (process)	Access to learning aids	Teachers using equipment, type and quantity of equipment available, etc.	Teacher survey
Students exposed to suitable learning aids (process)	Use of learning aids in classes	Amount, type and quality of learning aids used	Classroom report
Teachers intend to deliver the programme (outcome)	Intent	Teachers' self-reported intent to teach astronomy	Teacher survey
Teachers deliver the programme (outcome)	Behaviour	Number of classes taught etc.	Classroom report
Learners inspired by astronomy (outcome)	Excitement	Interest in astronomy, motivation to learn about astronomy	Primary school survey
	Superordinate group identification	Awareness, identification, absorption	Primary and secondary student surveys
Secondary school learners interested in science (outcome)	STEM activation	Interest in STEM, belief in importance of STEM, motivation to study STEM	Secondary student survey

Table 4: Surveys for children participating in astronomy outreach

Primary school student survey (for children aged 8 to 12 years)	
1	I find astronomy interesting
2	I like astronomy
3	I sometimes cannot stop thinking about the things we are taught in astronomy
4	I am excited to talk to my friends and family about what I learn in astronomy
5	I would like to read more about astronomy on my own, outside of class
6	When I look at the sky, I forget about the bad things that worry me
7	Learning about astronomy has shown me that I am part of a very big world
8	Learning about astronomy has opened my mind to other people who share my interest
Secondary school student survey (for children aged 13 to 18 years)	
1	I find science interesting
2	I like science
3	Learning science is important for understanding my place in the world
4	Learning science has changed my ideas about how the world works
5	Learning science will help me get a good job
6	I want to learn more about science
7	Learning about science showed me that I live in a very big world
8	Learning science has opened my mind to other people who share my interest
9	Learning about astronomy in science class was exciting

Response ranges from 1 (strongly disagree) to 4 (strongly agree), except for Item 9.

that they focused too much on specific STEM areas such as science or mathematics,^{44,45} making them more suitable for the evaluation of structured programmes solely focused on these areas of STEM activation. For both primary and secondary school children, we struggled to identify items that captured identification with a superordinate group at the level of universal citizenship, rather than, for example, superordinate groups at the level of nationality or ethnicity.⁴⁶ Given these challenges, after much discussion and research, we developed the surveys in Table 4 for primary school students into those for secondary school students.

Programme theory for the Public Outreach Programme

The main goal of this programme is to inspire, entertain and introduce adults to the accessible science of astronomy. The programme theory is shown in Figure 4.

The assumptions of the OAD underlying this programme are that the public will appreciate astronomy if this programme is delivered by

competent presenters. Here data monitoring is concerned with coverage of the public, the quality of the public presentation, and the attitudes of the audience towards astronomy and science (Table 5).

The items for the participant survey for this programme are presented in Table 6.

Interpreting outcomes

The surveys and instruments presented in this paper are intended to measure the extent to which the OAD's astronomy outreach activities are bringing about desirable changes in programme beneficiaries. Because these changes are seen as stepping stones towards the OAD's long-term strategic goal of astronomy for development, monitoring indicators of progress towards these benchmark intermediate outcomes is important. But how should the OAD interpret the monitoring data? And how should data be used more generally to guide programme improvement?

In programme evaluation, failure of a programme to achieve its outcomes is usually taken to imply one of two things: implementation

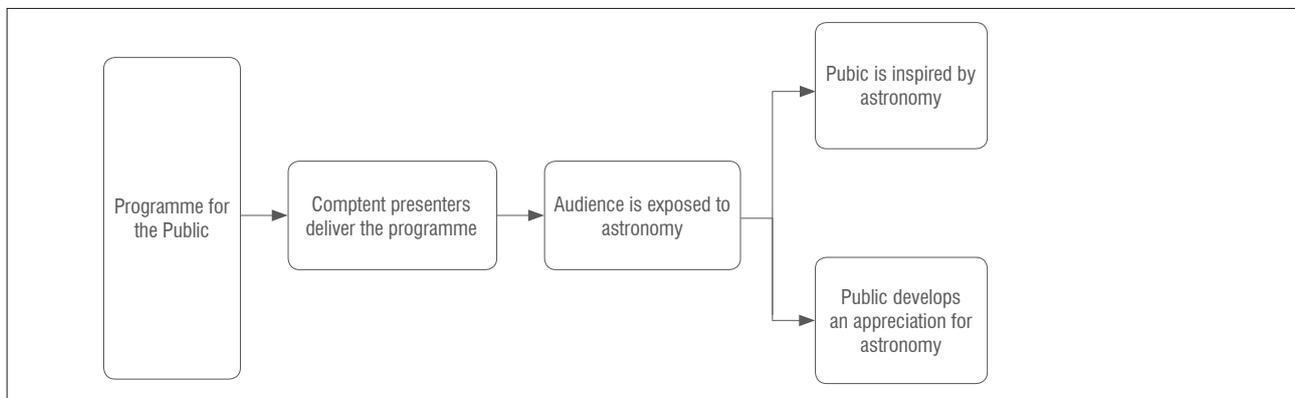


Figure 4: Programme theory for the Programme for the Public.

Table 5: Process and outcome indicators and measures for the Programme for the Public

Process/Outcome	Indicators	Measures	Data sources
Competent presenters deliver programme (process)	Amount, type and calibre of lecturers and lectures	Lecture outputs, audience lecturer ratings, etc.	Lecturer and lecture reports, participant survey
Public exposed to astronomy (process)	Use of small telescopes/planetariums	Proportion of lectures in which a small telescope or planetarium is used	Lecture report
Public develops an appreciation of astronomy (outcome)	Astronomy activation	Participants' self-reported change in astronomy activation	Participant survey

Table 6: Public participant survey for public's attitude towards astronomy and an astronomy lecture

1	The course presenter was enthusiastic about astronomy
2	The course presenter demonstrated a high level of astronomy expertise
3	This presentation has increased my interest in astronomy
4	I can't stop thinking about astronomy
5	I want to learn more about astronomy
6	I think astronomy is useful to help me understand science
7	Astronomy is relevant to what I experience in my life
8	I think astronomy is useful to make the world a better place

Response ranges from 1 (strongly disagree) to 4 (strongly agree).

failure or theory failure. Corrective action taken by the programme will vary considerably depending on the type of failure that is implicated. As Rossi et al.⁷ explain, implementation failure usually suggests that there has been poor service delivery on the part of the programme, and/or inadequate service utilisation on the part of the participants. Shortcomings in service delivery might be because activities were not delivered at the right intensity, resources were insufficient, or the amount, type and quality of interventions were inadequate. Service utilisation failure on the other hand might mean that beneficiaries have not responded to the programme in the manner intended, or that the wrong beneficiaries were targeted in the first place. The programme may also have had insufficient reach. Whereas corrective action in the former instance would usually centre efforts to improve service delivery, corrective action for a service utilisation failure is likely to focus on effort to improve the coverage, targeting and uptake of outreach programmes.

The second potential reason for a failure to achieve outcomes might be that the programme theory itself is implausible. Thus, if an implementation failure is not suggested by the M&E data, a theory failure must be considered likely. Even a very well delivered programme with an implausible theory can never be expected to bring about outcomes – because the logic linking actions to outcomes is fundamentally flawed. For us, the implications of a theory failure are potentially far more significant than those for an implementation failure, because in the case of theory failure, the strategic goals of the programme in relation to project activities might need to be reassessed and potentially revised.

Understanding process

From the above it is apparent that understanding process is critical to interpreting outcomes, because a good understanding of service delivery and utilisation can aid in distinguishing between a theory and implementation failure. However, our review of the astronomy outreach programme evaluation literature shows that indicators of implementation are seldom reported, making failure to achieve outcomes difficult to interpret. In the M&E framework presented here, we were therefore careful to incorporate process monitoring indicators and measures in an attempt to address this oversight. The fixed indicators and measures outlined in our M&E framework are, however, quite limited in nature and may occasionally need to be supplemented by more detailed process evaluations. Process evaluations, while similar to process monitoring in their scope and application, are typically tailored to specific projects and structured around key evaluation questions.

The next challenge: Measuring impact

In this article we have outlined an approach for monitoring and evaluating changes in the status of programme implementation and outcomes over time. Using this framework, regular measures taken over time are likely to provide valuable data suitable for formative programme improvement. The first priority of the OAD is therefore to test and apply the M&E framework to its global project sites. The objective will be to acquire as much quality data as possible and receive useful feedback on the framework's usability and relevance. This process of testing and revision can be tracked on the OAD website, from where the latest M&E framework and tools can also be downloaded.³⁵

The framework does not, however, allow for establishing a programme effect or the 'proportion of an outcome change that can be attributed uniquely to the programme as opposed to the influence of some other factor'^{16(p.206)}. In order to make this assessment, an impact evaluation with some form of control group would be needed. While the OAD has committed to reporting on the impact of their programmes for the period 2010–2020, a move towards more rigorous impact evaluation approaches for 'astronomy for development' will pose considerable challenges – many of which we have highlighted in this paper. In order to justify such a study, we would need to carefully consider their reasons for wanting to establish impact, and the way in which impact data would ultimately be used. Moreover, we would need to acknowledge that impact is unlikely to be achieved if there is implementation or theory failure.

In this paper, we have hoped to illustrate how the careful M&E of a programme's implementation and intermediate outcomes can greatly

assist in determining the plausibility of the programme theory. Without a plausible programme theory, attempting to gauge by empirical means as to whether a programme has brought about desirable long-term impacts is unlikely to prove a useful exercise. And if implementation is flawed, even short-term outcomes are unlikely to be realised. For these reasons, focusing on operationalising a simple M&E framework such as the one presented in this paper has numerous advantages. Should the framework indicate that the OAD's outreach initiatives are both theoretically and operationally sound, the next challenge would be to progress towards a more rigorous assessment of impact.

Acknowledgements

The OAD is a joint venture between the IAU and the South African National Research Foundation (NRF) with support of the Department of Science and Technology. We thank the South African Astronomical Observatory, a facility of the NRF, for hosting and facilitating the OAD.

Authors' contributions

S.C. wrote the manuscript and was responsible for the evaluation design. L.C. is a PhD student working with the OAD; she provided comments on the manuscript, wrote key sections on the role of astronomy in science, technology and development, and gave input on the evaluation design. J-C.M. is the programme manager for IAU OAD; he gave comments on the manuscript, gave input on the evaluation design and supervised L.C. K.G. is the OAD director; he gave comments on the manuscript, wrote key sections on the mandate and organisational structure of the OAD, contributed to the evaluation design, provided overall project oversight, and supervised L.C. J.L-P. is the corresponding author and project leader; she wrote the first outline of the manuscript, gave input on the subsequent drafts, was responsible for the evaluation design, and supervised S.C.

References

1. Shaffer PM, Whitney VW, Wills L, Senghore F, Matthews MJ. Measuring inspiration: Planning for NASA education's performance measurement and evaluation. In: 27th Annual Conference of the American Evaluation Association; 2013 Oct 16–18; Washington DC, USA. Washington DC: American Evaluation Association; 2013. p. 246–255.
2. International Astronomical Union (IAU). Astronomy for development: Building from the IYA 2009 – Strategic plan for 2012–2020. Cape Town: IAU; 2010.
3. Rosenberg M, Russo P. Why is astronomy important? International Astronomical Union 2013. Available from: http://www.iau.org/public/themes/why_is_astronomy_important/
4. Schilizzi RT, Dewdney PEF, Lazio JW. The Square Kilometre Array. Ground-based and airborne telescopes II. Proc SPIE. 2008;7012(11):1–13.
5. Promoting Africa European Research Infrastructure Partnerships (PAERIP). An analysis of the socio-economic impact of African-European research infrastructure cooperation. PAERIP; 2012. Available from: <http://www.paerip.org/sites/www.paerip.org/files/PAERIP%20report%203.3%2015092012.pdf>
6. Poh L. Research policy: How to build science capacity. Nature. 2012;490(7420):331–334. <http://dx.doi.org/10.1038/490331a>
7. Rossi PH, Lipsey MW, Freeman HE. Evaluation: A systematic approach. 7th ed. Thousand Oaks, CA: Sage Publications; 2004. Lelliott A, Rollnick M. Big ideas: A review of astronomy education research 1974–2008. Int J Sci Educ. 2009;32(13):1771–1799.
8. Diakidoy I-AN, Kendeou P. Facilitating conceptual change in astronomy: A comparison of the effectiveness of two instructional approaches. Learn Instruct. 2001;11(1):1–20. [http://dx.doi.org/10.1016/S0959-4752\(00\)00011-6](http://dx.doi.org/10.1016/S0959-4752(00)00011-6)
9. Hannust T, Kikas E. Children's knowledge of astronomy and its change in the course of learning. Early Child Res Quart. 2007;22(1):89–104. <http://dx.doi.org/10.1016/j.ecresq.2006.11.001>
10. Stroud N, Groome M, Connolly R, Sheppard K. Toward a methodology for informal astronomy education research. Astron Edu Rev. 2006;5(2):146–158. <http://dx.doi.org/10.3847/AER2006023>

11. Malyn-Smith J, Cedrone D. A program director's guide to evaluating STEM education programs: Lessons learned from local, state, and national initiatives [document on the Internet]. c2013 [cited 2014 May 27]. Available from: http://stelar.edc.org/sites/stelar.edc.org/files/A_Program_Directors_Guide_to_Evaluating_STEM_Education_Programs_links_updated.pdf
12. Frechtling JA. User-friendly handbook for project evaluation: Science Mathematics, Engineering and Technology education. Arlington, VA: The National Science Foundation; 1993.
13. Osborne J, Simon S, Collins S. Attitudes towards science: A review of the literature and its implications. *Int J Sci Educ.* 2003;25(9):1049–1079. <http://dx.doi.org/10.1080/0950069032000032199>
14. George-Jackson CE, Rincon B. Increasing sustainability of stem intervention programs through evaluation [document on the Internet]. c2013 [cited 2014 May 27]. Available from: <http://rube.asq.org/edu/2012/02/engineering/increasing-sustainability-of-stem-intervention-programs-through-evaluation.pdf>.
15. United Kingdom Museums, Libraries and Archives Council (MLA). Generic learning outcomes [homepage on the Internet]. c2014 [cited 2014 Dec 08]. Available from: <http://inspiringlearningforall.gov.uk>
16. Ciotti JI. Museums and planetariums: Bridging the gap between Hawaiian culture and astronomy through informal education – A case study [document on the Internet]. c2010 [cited 2014 June 26]. Available from: <http://forumonpublicpolicy.com/spring2010.vol2010/spring2010archive/ciotti.pdf>.
17. Arisa E, Fabregat J, Ros RM. EU-UNAWA Project: Inspiring every child with our wonderful cosmos. *Highlights Span Astrophysics VII.* 2013;1:959–961.
18. Burtnyk K. Impact of observatory visitor centres on the public's understanding of astronomy. *Publ Astron Soc Aust.* 2000;17:275–281. <http://dx.doi.org/10.1071/AS00043>
19. Prather EE, Cormier S, Wallace CS, Lintott C, Jordan Raddick M, Smith A. Measuring the conceptual understandings of citizen scientists participating in Zooniverse projects: A first approach. *Astron Edu Rev.* 2013;12(1):010109. <http://dx.doi.org/10.3847/AER2013002>
20. Rhodes H, Neishi K, Velez M, Mendez J, Martinez A. The Sol pilot national evaluation: Findings and recommendations. Cambridge, MA: Abt Associates; 2011.
21. Friedman AJ, editor. Framework for evaluating impacts of informal science education projects: Report from a National Science Foundation workshop. Arlington, VA: National Science Foundation; 2008. Available from: http://www.aura-astronomy.org/news/EPO/eval_framework.pdf
22. Jarman R, McAleese L. Physics for the star-gazer: Pupils' attitudes to astronomy in the Northern Ireland science curriculum. *Phys Educ.* 1999;31(4):223–226. <http://dx.doi.org/10.1088/0031-9120/31/4/020>
23. Jarvis T, Pell A. Effect of the challenger experience on elementary children's attitudes to science. *J Res Sci Teach.* 2002;39(10):979–1000. <http://dx.doi.org/10.1002/tea.10055>
24. Jarvis T, Pell A. Factors influencing elementary school children's attitudes toward science before, during, and after a visit to the UK National Space Centre. *J Res Sci Teach.* 2005;42(1):53–83. <http://dx.doi.org/10.1002/tea.20045>
25. Zeilik M, Schau C, Mattern N. Conceptual astronomy. II. Replicating conceptual gains, probing attitude changes across three semesters. *Am J Phys.* 1999;67(10):923–927. <http://dx.doi.org/10.1119/1.19151>
26. Ucar S, Demircioglu T. Changes in preservice teacher attitudes toward astronomy within a semester-long astronomy instruction and four-year-long teacher training programme. *J Sci Educ Teach.* 2011;20(1):65–73. <http://dx.doi.org/10.1007/s10956-010-9234-7>
27. Falk JH, Koran JJ, Dierking LD. The things of science: Assessing the learning potential of science museums. *Sci Educ.* 1986;70(5):503–508. <http://dx.doi.org/10.1002/sce.3730700504>
28. Şentürk E, Özdemir ÖF. The effect of science centres on students' attitudes towards science. *Int J Sci Educ.* 2012;4(1):1–24. <http://dx.doi.org/10.1080/21548455.2012.726754>
29. Prather EE, Slater TF, Adams JP, Bailey JM, Jones LV, Dostal JA. Research on a lecture-tutorial approach to teaching introductory astronomy for non-science majors. *Astron Edu Rev.* 2004;3(2):122–123. <http://dx.doi.org/10.3847/AER2004019>
30. Coryn CLS, Noakes LA, Westine CD, Schröter DC. A systematic review of theory-driven evaluation practice from 1990 to 2009. *Am J Eval.* 2011;32(2):199–226. <http://dx.doi.org/10.1177/1098214010389321>
31. Weiss CH. Theory-based evaluation: Past, present, and future. *New Dir Eval.* 1997;1997(76):41–55. <http://dx.doi.org/10.1002/ev.1086>
32. Chen HT. Theory-driven evaluations. New York: Sage; 1990.
33. Bickman L. Using program theory in evaluation. In: Bickman L, editor. *New directions for program evaluation*, no 33. San Francisco, CA: Jossey-Bass; 1987. p. 5–18.
34. Office of Astronomy for Development (OAD). Monitoring and evaluation framework of the OAD [homepage on the Internet]. c2014 [cited 2015 Apr 27]. Available from: <http://www.astro4dev.org/funded-projects-2014/me-framework/>
35. Atwood VA, Atwood RK. Preservice elementary teachers' conceptions of what causes night and day. *School Sci Math.* 1995;95(6):290–294. <http://dx.doi.org/10.1111/j.1949-8594.1995.tb15785.x>
36. Mant J. A survey of British primary school teachers' understanding of the Earth's place in the universe. *Educ Res.* 1995;37(1):3–19. <http://dx.doi.org/10.1080/0013188950370101>
37. Greene JC, DeStefano L, Burgon H, Hall J. An educative, values-engaged approach to evaluating STEM educational programs. *New Dir Eval.* 2006;109:53–71. <http://dx.doi.org/10.1002/ev.178>
38. Zeyer A, Çetin-Dindar A, Zain M, Jurišević M, Devetak I, Odermatt F. Systemizing: A cross-cultural constant for motivation to learn science. *J Res Sci Teach.* 2014;50(9):1047–1067. <http://dx.doi.org/10.1002/tea.21101>
39. Clewell BC, Cosentino de Cohen C, Tsui L, Deterding N. Revitalizing the nation's talent pool in STEM: Science, technology, engineering and math [document on the Internet]. c2006 [cited 2014 May 27]. Available from: http://www.urban.org/UploadedPDF/311299_revitalizing_stem.pdf
40. Tsui L. Effective strategies to increase diversity in STEM fields: A review of the research literature. *J Negro Educ.* 2007;76(4):555–581.
41. Bektaşlı B. The effect of media on preservice science teachers attitudes towards astronomy and achievement in astronomy class. *Turk Online J Ed Tech.* 2013;12(1):139–146.
42. Hartry A, Snow J. A toolkit for evaluating small-scale science, technology, engineering, and mathematics (STEM) programs: Big mistake/bad idea? Or a way to ensure small organizations have access to high-quality evaluations? In: 27th Annual Conference of the American Evaluation Association; 2013 Oct 16–18; Washington DC, USA. Washington DC: American Evaluation Association; 2013. p. 142–160.
43. Germann PJ. Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. *J Res Sci Teach.* 1988;25(8):689–703. <http://dx.doi.org/10.1002/tea.3660250807>
44. Kind P, Jones K, Barmby P. Developing attitudes towards science measures. *Int J Sci Educ.* 2007;29(7):871–893. <http://dx.doi.org/10.1080/09500690600909091>
45. Stone CH, Crisp RJ. Superordinate and subgroup identification as predictors of intergroup evaluation in common ingroup contexts. *Group Process Intergr Relat.* 2007;10(4):493–513. <http://dx.doi.org/10.1177/1368430207081537>

