

Large-scale international assessments of learning outcomes: balancing the interests of multiple stakeholders

Guirong Li, Irina Shcheglova, Ashutosh Bhuradia, Yanyan Li, Prashant Loyalka, Olivia Zhou, Shangfeng Hu, Ningning Yu, Liping Ma, Fei Guo & Igor Chirikov

To cite this article: Guirong Li, Irina Shcheglova, Ashutosh Bhuradia, Yanyan Li, Prashant Loyalka, Olivia Zhou, Shangfeng Hu, Ningning Yu, Liping Ma, Fei Guo & Igor Chirikov (2021) Large-scale international assessments of learning outcomes: balancing the interests of multiple stakeholders, *Journal of Higher Education Policy and Management*, 43:2, 198-213, DOI: [10.1080/1360080X.2020.1767327](https://doi.org/10.1080/1360080X.2020.1767327)

To link to this article: <https://doi.org/10.1080/1360080X.2020.1767327>



Published online: 25 May 2020.



Submit your article to this journal [↗](#)



Article views: 562



View related articles [↗](#)



View Crossmark data [↗](#)



Large-scale international assessments of learning outcomes: balancing the interests of multiple stakeholders

Guirong Li^{a,b}, Irina Shcheglova^c, Ashutosh Bhuradia^d, Yanyan Li^{a,b,e},
Prashant Loyalka^d, Olivia Zhou^d, Shangfeng Hu^f, Ningning Yu^g, Liping Ma^h, Fei Guoⁱ
and Igor Chirikov^{c,j}

^aSchool of Education, Henan University, Kaifeng, China; ^bInternational Center for Action Research on Education, Henan University, Kaifeng, China; ^cInstitute of Education, National Research University Higher School of Economics, Moscow, Russia; ^dGraduate School of Education, Stanford University, Stanford, CA, USA; ^eRural Education Action Program, Stanford University, Stanford, CA, USA; ^fStudent Affairs Office, Sichuan Normal University, Chengdu, Sichuan, China; ^gInstitute of Higher Education Research, University of Jinan, Jinan, Shandong, China; ^hGraduate School of Education, Peking University, Beijing, China; ⁱInstitute of Education, Tsinghua University, Beijing, China; ^jCenter for Studies in Higher Education, UC Berkeley, Berkeley, CA, USA

ABSTRACT

The demand for large-scale assessments in higher education, especially at an international scale, is growing. A major challenge of conducting these assessments, however, is that they require understanding and balancing the interests of multiple stakeholders (government officials, university administrators, and students) and also overcoming potential unwillingness of these stakeholders to participate. In this paper, we take the experience of the Study of Undergraduate Performance (SUPER) in conducting a large-scale international assessment as a case study. We discuss ways in which we mitigated perceived risks, built trust, and provided incentives to ensure the successful engagement of stakeholders during the study's implementation.

KEYWORDS

Higher education; large-scale assessment; skills; China; India; Russia

Introduction

Both developed and developing countries set goals to increase the number of its highly skilled college graduates by widening access to and improving the quality of higher education (Altbach, Reisberg, & Rumbley, 2009). Policymakers are increasingly relying on a variety of efforts to improve the quality of higher education. These efforts at quality assurance include assessments, accreditation, and audits (Eaton, 2003; Woodhouse, 1998). They are indicative of a general trend – a shift in focus from input-based measures of student success such as student-teacher ratios and teacher qualifications to outcome-based measures: student learning and employment outcomes (Kardanova et al., 2016; Loyalka et al., 2019).

While some outcome-based measures of quality such as student self-reported satisfaction or employability are becoming more salient in accreditation programmes (ABET 2015; Brahimi, Sarirete, & Ibrahim, 2016), direct measures of student learning outcomes

are rarely used at scale in quality assurance. As such, policymakers are unable fully to assess the quality of higher education institutions and hold them accountable for student learning. Furthermore, without measures of learning outcomes over time or across countries, policymakers are unable to contextualise the 'health' of their higher education institutions and systems in a globally competitive environment. The lack of such benchmarks at both national and international levels has increased demand for large-scale student assessments in higher education (Alexander, 2000; Steedle, 2012).

A major challenge of conducting large-scale assessments in higher education, however, is that they require engaging multiple stakeholders (government officials, private donors, university administrators, students, and academic staff) and balancing their varied interests (Jongbloed, Enders, & Salerno, 2008; Lyytinen et al., 2017). In most cases large-scale assessments also require overcoming the unwillingness of key stakeholders to participate. The stakeholders will give permission or choose to participate in the large-scale assessments only if the perceived benefits outweigh potential risks emerging when full results are known to other stakeholders or made public.

To illustrate, a potential benefit of conducting large-scale assessments is that they claim they can provide useful information about skill gaps among college graduates (Ewell, 2012; Shahjahan & Torres, 2013). Stakeholders can use this information to make resource allocation decisions. For example, if students at an institution or in a country are behind in critical thinking skills, stakeholders can decide to invest resources in pedagogy that help students learn such skills.

On the flipside, however, conducting large-scale assessments could also be risky if they promote transparency to such an extent that it compromises the interests of stakeholders. For example, institutions might fear that unfavourable news about their performance in a large-scale assessment could discourage prospective students from enrolling in their programmes. Stakeholders might also be mistrustful of large-scale assessments, especially in an international context, if they think their performance could undermine their reputation on the global stage. It is not surprising then that large-scale international assessments in the past have been hampered by disengagement among stakeholders (Lennon, Fukahori, & Edwards, 2018).

Despite their important role, past research has seldom explored how stakeholders can best be engaged to participate in large-scale assessments in higher education. To the best of our knowledge, few studies have systemically outlined stakeholder interests and concerns. Furthermore, few studies have examined ways to address these interests and concerns.

Our study therefore has two goals. First, we seek to highlight the interests and concerns of three major stakeholders in higher education assessment – government officials, university administrators, and students. Second, we propose strategies to address and balance these interests and concerns. We will rely on the stakeholder theory (Donaldson & Preston, 1995; Freeman, 1984; Reynolds, Schultz, & Hekman, 2006) as a conceptual framework to identify the stakeholders, describe their needs, interests and influences and the relationships with these stakeholders that will lead to the success in higher education assessment. In particular, we consider the provision of incentives, efforts to mitigate perceived risks, the means to build trust among varied stakeholders and incorporate stakeholders in the decision-making process.

We use the experience of the SUPER programme in three countries – China, India, and Russia – as a case study on engaging multiple stakeholders in large-scale assessments.

The paper is based on authors' experiences with implementing this programme. Besides the OECD's AHELO programme, the SUPER programme is the only other large-scale international assessment in higher education of which we are aware. The international aspect of the programme affords additional insights into balancing the interests and concerns of multiple stakeholders. It also offers particular challenges in addressing these interests and concerns, while maintaining the international comparability of the assessment results.

Background

The SUPER programme

The SUPER programme is organised by researchers at Stanford University in collaboration with Educational Testing Services (ETS) and university-based research teams in China, Russia and India. The research team includes higher education researchers, psychometricians as well as experts in engineering education. A major aim of the SUPER programme is to assess academic (such as maths, science, and major-specific skills) and higher-order thinking skill (such as critical thinking and creativity) levels and gains among computer science and electrical engineering students in undergraduate programmes in China, Russia, and India. Computer Science (CS) and Electrical Engineering (EE) majors were selected because they produce graduates who often compete for jobs internationally and thus it is important to understand relative student learning across countries (Kardanova et al., 2016). Another objective of the SUPER programme is to understand what factors (e.g. student, peer, institutional, academic staff) help students make those gains. Finally, the programme seeks to examine skills and the factors leading to skill gains for students in different types of institutions (e.g. elite and non-elite; public and private) and from different demographic subgroups (e.g. male and female students).

The SUPER programme started in 2014 and included three stages:

During the pilot stage, the research team developed and piloted valid and reliable assessment instruments that can be used to assess and compare skill levels and gains in mathematics and physics for electrical engineering and computer science students. The development of the assessment instruments included cross-national expert evaluations of content areas and test items for mathematics and physics, the analysis of content and construct validity of the instruments, the pilot student survey and the analysis of the psychometric quality of the instruments: dimensionality, reliability, cross-national comparability (see Kardanova et al. (2016) for more details on the instruments development). The team also translated and adapted a critical thinking test and a quantitative literacy test that were previously developed and piloted by ETS (Gu et al., 2018; Liu et al., 2018; Shaw et al., 2019).

Next, during the baseline stage the research team drew, assessed and surveyed nationally representative samples of institutions and students in each country. In China, Russia and India the research team first identified all undergraduate CS and EE programmes with course requirements that were comparable across countries. It then randomly sampled institutions that offered these programmes. In China, it

sampled 36 institutions in total – 6 institutions each from 6 representative provinces. In India and Russia, it sampled 50 and 34 institutions respectively based on a stratified national random sample. At each sampled institution, the team randomly selected an administrative unit within CS and EE programmes and included all first and third-year students in those units. The overall sample consisted of ~8,000 students in China, ~4,500 students in Russia and ~18,000 students in India (grade 1 and grade 3 undergraduate students majoring in EE and CS). Next, the team assessed these students on academic and higher-order thinking skills – specifically, it randomly administered academic skills exams to half the students and higher-order thinking exams to the remaining half. The team also collected contextual survey data on students, professors, and administrators.

During the endline stage, the research team returned to the same institutions in China, Russia and India two years later to track down the study participants and invite them to take the second set of assessments. Grade 2 students took academic skills exams and higher-order thinking exams. Grade 4 students took a major-specific test in either computer science (developed by ETS) or electrical engineering (developed by our project team) as well as tests of critical thinking, quantitative literacy, and creativity. As in the baseline stage, the team collected contextual survey data on students and professors.

When the research team designed and implemented the SUPER programme, it took into account the experience of other international assessment programmes and, specifically, the AHELO feasibility study (Ashwin, 2015; Ewell, 2012; Lennon et al., 2018; Shahjahan & Torres, 2013; Tremblay, Lalancette, & Roseveare, 2012). Similar to the AHELO, the SUPER programme focuses on the international assessments of discipline-specific and generic skills in undergraduate engineering programmes. However, there are important differences between the AHELO and the SUPER programme in the scope, organisation and the design of the study.

First, the AHELO programme had a much wider scope both in terms of the assessed academic majors (engineering, economics) and the number of countries involved in the study. By contrast, the SUPER programme focuses on two related majors of computer science and electrical engineering. The smaller scale allowed for a more careful selection of the content for the assessment of academic and higher-order thinking skills that reflected the differences in countries' curricular requirements. It also helped to better coordinate and synchronise the work of the research teams in each country. The smaller scale allowed the SUPER programme to be funded through research grants and alleviated the pressure on the national governments to allocate funding. The latter was a significant challenge during the implementation of the AHELO feasibility study (Tremblay et al., 2012).

Second, there are a number of differences in the design of the study. The AHELO measured skill levels of engineering students at the end of university and it was unable to separate whether students learned skills during their time at university or before that. By contrast, the longitudinal design of the SUPER programme allows to assess the actual skill gains (how much students learn over time) and compare these skill gains within and across countries. Unlike the AHELO, the SUPER programme uses nationally representative (random) samples of institutions and students within institutions. The SUPER-programme also collected rich contextual information about students, their professors and departments that allowed to use causal research designs to understand the impact of different factors on student

learning. This information was particularly important to the policy-makers in all participating countries because it helps develop evidence-based interventions.

Recent higher education expansion and the need for quality assurance in China, Russia and India

Higher education in China, Russia, and India has rapidly expanded in the last two decades. Privatisation, to varying degrees, has been an important driver of expansion in these countries (Carnoy et al., 2013). In China, privatisation and marketisation allowed for the establishment of private and vocational colleges that increased enrolment in higher education and consequently helped address the demand for skilled labour in the economy (Wang, 2014). Quality assurance in China is coordinated by the Ministry of Education (MoE) and is often criticised for being too bureaucratic and not focusing on student learning (Jiang, 2015).

In Russia, newly established private universities as well as former Soviet public higher education institutions that were given permission by the government to open new demand-driven academic programmes (and charge tuition directly from students) have contributed to growth in higher education (Froumin, Kouzminov, & Semyonov, 2014). Quality assurance system in Russia is also coordinated by the government and focuses on the formal indicators (e.g. infrastructure, instructor qualifications) rather than on student learning.

In India, higher education expansion has coincided with the passing of the Private Universities Act in 2003, which greatly enhanced the participation of private institutions (Varghese & Malik, 2015). Now, a majority of college students in India are enrolled in private institutions (Marmolejo & Beteille, 2017). While rapid growth has improved access to higher education for millions of youth in these countries, it has also increased the burden on policymakers to ensure institutions meet certain standards of quality and accountability. In India, 15 professional councils regulate quality in different sectors of higher education – for example, the All India Council of Technical Education (AICTE) for Engineering and Computer Science. India appears to have a large and well-developed quality assurance system but it is also criticised for being rigid and burdensome (Kapur & Perry, 2015).

All three countries face many challenges associated with implementing quality assurance efforts in the era of massified higher education sector. First, many indicators that are used in quality assurance do not assess actual student learning outcomes as they focus mostly on customer protection against imitation and poor-quality education. Second, the expansion of higher education seems to have outpaced the capacity of governments in these countries to monitor quality at scale, which has arguably created a demand for large-scale assessments of student learning outcomes.

Stakeholder identification

As discussed, large-scale assessments can potentially offer a practical solution to quality assurance and accountability issues in developed and developing countries. However, international large-scale assessments also require understanding the interests and concerns of different stakeholders in each country, and devising strategies to address these. We utilise stakeholder theory (Donaldson & Preston, 1995; Freeman, 1984) to identify stakeholders of the SUPER programme, describe their interests and concerns, and

analyse strategies addressing these interests and concerns. Stakeholder theory had been extensively used in the studies of quality assurance and assessment in higher education (Jongbloed et al., 2008; Benneworth & Jongbloed, 2010; Leisyte & Westerheijden, 2014; Lyytinen et al., 2017). Following these studies, we will rely on the classical definition of stakeholder as the one who ‘can affect or is affected by the achievement of the organisation’s objectives’ (Freeman, 1984, p. 46). Stakeholder theory has been a useful heuristic to analyse the environment of organisations, programmes or initiatives.

One of the most common strategies to identify stakeholders is to apply a three-factor ‘stakeholder salience’ model (Mitchell, Agle, & Wood, 1997). This model uses three criteria to identify different levels of stakeholder engagement: power, legitimacy, and urgency. Power refers to the ability of stakeholder to influence the decisions employing coercive, utilitarian or normative means (Mitchell et al., 1997, p. 865–866). Legitimacy emphasises socially accepted behaviours that allow to establish authority of the stakeholder (Mitchell et al., 1997, p. 866–867). Urgency refers to how quickly an organisation should respond to a request of a stakeholder (Mitchell et al., 1997, p. 867–868).

When applied to the SUPER programme, power refers to the ability of stakeholders to influence the content and organisational design of the programme, provide or withdraw resources to support it. Legitimacy is a shared expectation of a stakeholder to officially participate in the programme and be able to have access to reports and other deliverables. Urgency highlights the ability to claim a call for an immediate attention in the implementation stage of the programme (providing questionnaire, logistical information, reports, etc.).

According to ‘stakeholder salience’ model (Mitchell et al., 1997), salience is determined by the cumulative number of stakeholder attributes – power, legitimacy, and urgency. Stakeholders that possess only one attribute are classified as ‘latent’, two attributes – as ‘expectant’, three – as ‘definitive’ stakeholders (Mitchell et al., 1997). Managers develop different strategies depending on the type of the stakeholder and balance their interests accordingly. In the following section, we will identify different types of stakeholders among policymakers, universities, and students as well as their interests and concerns in China, Russia and India.

Policymakers

In China, the SUPER programme team approached the key policy-level stakeholder, the Ministry of Education (MoE), during the pre-implementation phase to seek their support and involvement. The MoE had been investing resources to improve engineering education – for instance, the Chinese government had been promoting ‘Excellent Engineers Cultivation Plan (2010–2020)’ – and therefore seemed like natural allies. Another rationale for having the MoE involved was simply operational – their official support would have ensured a relatively smooth implementation of the programme.

When approached by the program team, the MoE were indeed enthusiastic and expressed interest in large-scale international assessments of student learning outcomes. However, they were also wary of lending official support. Their cautious approach may have stemmed from uncertainty about what the programme might reveal – specifically, about how students in China were progressing in college and how they were performing compared to their peers in other parts of the world. Their wait-and-watch approach may

have been prudent from a political perspective. In terms of stakeholder salience model, MoE was a latent stakeholder of the programme in China, possessing power attribute but not legitimacy and urgency attributes.

In Russia, the main policy-level stakeholder was the Ministry of Education and Science (MoES). The MoES saw this programme as a path to expanding quality assurance instruments and including the evaluation of academic and higher-order thinking skills in its Monitoring process. Three aspects of the programme were particularly appealing to the MoES. First, the programme focused on engineering education. The quality of engineering education is a top priority of the MoES due to a lack of interest in STEM disciplines among school graduates and concerns from employers about the quality of engineering education. Second, the SUPER programme was intent on benchmarking learning outcomes internationally. Since 2012, international comparisons and global rankings in higher education have become important components of strategic initiatives in higher education at the national level in Russia. Third, the SUPER programme included assessments not only of major-specific skills but also higher-order thinking skills. Higher-order thinking skills were important for the MoES because they were emphasised in the new Federal Standard of higher education and the MoES wanted to create a baseline for assessment of the current level of critical thinking.

The MoES conveyed very few concerns regarding the programme and were in fact well-positioned and resourced to support it. They played an important role in providing data for sampling and later in encouraging universities to participate. They also emphasised several topics that were of special interest to them and urged the team to explore these in the programme (e.g. factors that affect critical thinking skills and the impact of teacher qualifications on student outcomes). MoES is, therefore, a 'definitive stakeholder' of the SUPER programme in Russia, possessing power, legitimacy and urgency characteristics.

In India, there were two stakeholders at the policy level: The All India Council for Technical Education (AICTE), which is a national-level regulatory body that oversees technical education in India; and the Ministry of Human Resource Development (MHRD). The AICTE has overseen an impressive expansion of engineering education in India in recent decades. However, there were some concerns within the AICTE's leadership that this expansion may have come, to some degree, at the cost of quality. At the same time, external stakeholders have questioned the AICTE's legitimacy. The Indian government, for instance, had contemplated merging the AICTE and the UGC to create a single regulator – Higher Education Evaluation and Regulation Authority (HEERA). However, the AICTE opposed the measure contending that it is in the process of reforming its current regulatory approach and can still add value to the higher education sector as an independent body.

In this context, the AICTE's interest in the SUPER programme is not surprising. For example, an important aspect of the programme is not only to assess academic skills (which are traditionally measured in India) but also critical thinking skills that employers tend to value highly. An understanding, especially of the latter, could be vital to improving labour market outcomes of engineering students in India thus ensuring the AICTE's legitimacy. Another aspect of the programme is to provide input on the factors (student, academic staff, peer, institutional) that affect skill development – this was ostensibly important to the AICTE from both quality assurance and resource allocation perspective.

The other policy-level stakeholder, the MHRD, plans and coordinates higher education policy in India. It works closely with regulatory bodies (such as the UGC and the AICTE) to develop India's human capital. The MHRD's interest in large-scale international assessments, specifically the SUPER programme, was in understanding how India's 'elite' colleges were performing compared with their counterparts in other countries. This knowledge would be valuable to the MHRD for developing world-class colleges and universities in India.

Both the AICTE and the MHRD were enthusiastic about the programme from the beginning and found that the programme's goals aligned with their own. Consequently, they provided funding and official support to conduct the programme in India. While they did not convey concerns, they were interested in specific policy questions that they thought were important for improving engineering education in India (e.g. they were concerned about the quality of teaching in Indian colleges and hoped that the programme could provide feedback on how to improve pedagogy). Finally, they were also interested in learning ways in which their local units could independently implement similar large-scale assessment programmes in future. AICTE has created a unit that was working together with the programme team on the SUPER programme and developing capacity for higher education assessments at the national level in the future.

In terms of stakeholder salience model AICTE was a 'definitive stakeholder' possessing power, legitimacy and urgency characteristics. MHRD was an 'expectant stakeholder' possessing power and legitimacy characteristics (in Mitchell et al. (1997) terms a stakeholder that has power and legitimacy is also called a 'dominant' stakeholder).

Universities

The SUPER programme surveyed students, academic staff, and administrators at a combination of 'elite' and 'non-elite' universities in each country – determined either by a special status and/or selectivity of admissions.

In China, the 36 participating universities, all came in with varied interests and concerns to participate in the programme. Some institutions expressed concerns about the quality of higher education in China and saw the programme as a means to address those concerns. Specifically, they believed the programme could create benchmarks in education research and policy that could accelerate the improvement of higher education in China. Others wanted to understand a broad range of skills their students were learning over time and how the skill levels (and gains) of their students stacked up against other institutions. Since in China MoE did not assist in encouraging institutions to participate in the SUPER programme the team established and sustained horizontal relations with the universities from the sample. Each university was separately invited by the research team to participate. In terms of the stakeholder salience model universities in China possessed both legitimacy and urgency characteristics as they required immediate attention and responses to their requests.

In Russia, all universities were invited to participate via the letter from the MoES. These universities were a heterogeneous group with different expectations and concerns. There was an important dimension that differentiated this group: elite and non-elite institutions. Elite institutions (National Research Universities) were less dependent on the MoES and more demanding in terms of the outcomes they hoped to receive from the

programme (e.g. having regular reports). Non-elite universities were less demanding, and some agreed to participate simply to win favour with MoES. Elite universities possessed legitimacy and urgency characteristics of stakeholders, while non-elite universities had only a legitimacy characteristic.

In India, universities were invited to participate via AICTE and MHRD. A handful of participating colleges were sceptical about the SUPER programme. The administrators at these colleges were concerned that the programme was another form of government interference in an already over-regulated system. On the other hand, many institutions were interested in improving quality and obtaining a better sense of how their students are progressing in college. For example, many private institutions had incentives to enhance quality and attract students to enrol into their programmes. This was especially true in India where the supply of engineering seats has outstripped their demand, and thus colleges are aggressively vying for a limited number of students. High involvement of AICTE and MHRD also helped to incentivise universities, but in terms of stakeholder salience model universities possessed only legitimacy characteristic within the SUPER programme.

Participating institutions in all countries, however, had some common concerns. First, they were worried that university-level performance data could be misused and used punitively by regulators or negatively influence other stakeholders such as students, parents, and staff. Second, they expressed concerns about having teams of external proctors administer the assessment – specifically, they were concerned that proctors might not administer the assessments in a fair and standardised manner. Finally, they were unsure if participating in the programme would be worth their time in the absence of tangible benefits.

Students

The SUPER programme team tested and surveyed over 60,000 students in China, Russia, and India. Understanding their interest and concerns was vital to getting the buy-in of each of these stakeholders. Students were interested in the assessment because it measured certain skills (e.g. creativity and critical thinking) that were not typically evaluated. Some students were also excited about participating in a study designed by researchers at top research institutions and were curious about what the assessments had to offer. However, much like university administrators, they too questioned the benefit of investing their time to participate in the programme since the assessment was not directly related to their performance at school. Students were invited to participate by coordinators at each university. They were provided with an informed consent form by enumerators explaining that they had an opportunity to opt out from the programme without any consequences for their studies. They were also ensured that their performance in the programme won't affect their grades and were encouraged to do their best on the tests.

In all three countries, students can be characterised as latent stakeholders possessing only the legitimacy characteristic. The major challenge for the programme team was to increase their involvement and motivate them to participate.

Table 1 summarises salience characteristics of the SUPER programme stakeholders. Comparative analysis shows that there was a unique stakeholder composition in each country. In China, there were no definitive stakeholders of the programme while in Russia and India there were MoES and AICTE, respectively. In addition, in India there was another dominant stakeholder – MHRD that possessed both power and legitimacy.

Table 1. Stakeholder salience in the SUPER programme.

	Power	Legitimacy	Urgency
China	MoE*	Universities** Students*	Universities**
Russia	MoES***	MoES*** Students* Elite Universities** Non-Elite Universities*	MoES*** Elite Universities**
India	AICTE*** MHRD**	AICTE*** MHRD** Students* Universities*	AICTE***

* – latent stakeholders

** – expectant stakeholders

*** – definitive stakeholders

Universities were characterised as expectant stakeholders in China and Russia (their elite subset only) but not in India. Students were latent stakeholders in all three countries. These different stakeholder compositions in China, Russia and India required different approaches in balancing interests of multiple stakeholders.

Addressing interests and concerns of stakeholders

Prioritising competing interests of stakeholders has always been a challenge for quality assurance programmes in higher education (Jongbloed et al., 2008; Lyytinen et al., 2017). According to the stakeholder management perspective, there are two approaches that managers and policymakers can employ to balance the interests of multiple stakeholders: a within-decision approach and an across-decision approach (Reynolds et al., 2006). Within-decision approach means that the interests of stakeholders are balanced within the bounds of every decision that is made by the manager. For example, the interests of all stakeholders could be balanced separately for every decision made about the design and implementation of the SUPER programme – the content areas of assessment, development of survey instruments, sampling, etc. On the contrary, across-decision approach allows to balance interests of stakeholders as a whole: for some decisions the interests of particular stakeholders are sacrificed but then these stakeholders are compensated on future decisions (Reynolds et al., 2006, p. 289). In the context of unequal stakeholder salience and limited resource divisibility across-decision approach lead to the higher instrumental value and more ethical results (Reynolds et al., 2006).

The SUPER programme team employed the across-decision approach in balancing interests of multiple stakeholders. The strategies accounted for the salience of stakeholders and aimed at increasing the value of the project for all of them. The strategies involved four types of activities: providing incentives, building trust, mitigating stakeholders' perceived risks and integration in the decision-making process. We tried to balance the interests of stakeholders across these four types of activities. The more salient the stakeholder, the higher the number of activities were used to address its concerns. For example, working with definitive stakeholders required not only providing incentives and building trust (as with latent and expectant stakeholders) but also including them in the decision-making process. The decisions for each country were

made by the research team at Stanford University in close consultations with the local research teams in each country. The decision-making process aimed at acquiring stakeholder buy-in without compromising the original research design and comparability across countries. This section summarises the strategies by the type of stakeholder: latent, expectant and definitive.

Latent stakeholders

Latent stakeholders of the programme possessed only one characteristics of the stakeholder salience model – legitimacy. These were students in all three countries, non-elite universities in Russia and universities in India. The major goal of the programme team was to provide meaningful incentives for them to participate.

Incentives were crucial for students in particular because an important goal of any assessment programme is to obtain a high response rate from respondents. One method of doing so is to provide payment to those who participate. However, payments (an incentive used in surveys and assessments) are not allowed in some research settings. During the piloting phase of the SUPER programme, the research team conducted extensive research on how to elicit student participation and engagement. The experiment aimed at identifying incentives that would stimulate students to take the exams more seriously and aim for higher results on the test. Students were randomly assigned to receive one of four different informational prompts before the test: (1) that they will be provided with a personal report of the test results, (2) that the test results will be used to compare their university with other universities in the country, (3) that the content area of the test they are taking is very important, and (4) no prompt. The results showed that providing students with the option of receiving individualised (and confidential) reports of their performance on the assessment was important for boosting student engagement (0.1 SD higher than no-prompt group).

Another category of latent stakeholders are universities in India and non-elite universities in Russia. In addition to the efforts of MoES in Russia and both AICTE and MHRD in India, which were crucial in securing their participation, the programme team also provided additional incentives for these universities. The team designed detailed reports for universities using institutional-level data. These reports included the comparison of student scores at a particular institution with the average scores of students nationally and internationally. This also had a positive effect on incentivising universities to participate.

Expectant stakeholders

Expectant stakeholders of the SUPER programme included universities in China, elite universities in Russia and the MHRD in India. Universities in China and elite universities in Russia possess legitimacy and urgency, while MHRD possesses power and legitimacy. In addition to providing incentives for the stakeholders, the programme team put emphasis on mitigating risks and building trust with these stakeholders.

A perceived risk for multiple stakeholders (e.g. institutions, administrators etc.) in each country was whether their performance on the assessment would impact them negatively if results were made public or available to other stakeholders. The programme

team addressed this concern by ensuring confidentiality at every level of the survey. In terms of reporting the results, all students and universities participated on the following conditions. First, student-level reports were confidential and only students themselves had access to their individual results but not universities or regulatory bodies. Second, university-level reports were kept confidential as well but shared directly and only with universities themselves.

Policymakers and university administrators were both concerned about the quality of external proctoring. To address this concern, the programme team designed a survey protocol that was common for all countries (e.g. almost identical scripts for introducing the programme and motivating stakeholders to participate in each country). While the enumerator teams were hired locally, they were all trained to administer the programme in a standardised manner across countries. This ensured consistency and comparability especially in an international context.

Strategies to mitigate perceived risk and providing incentives would have been less effective without first building trust with stakeholders such that they felt the programme team would follow through on assurances. In this regard, the programme's team was comprised of local collaborators who were involved from the very beginning (e.g. the pilot phase) and who had a nuanced understanding of the expectations of different stakeholders and were very responsive to their interest and concerns. Trust-building particularly paid off in Russia. The leadership at the MoES, that had an important role in encouraging institutes to participate, changed while the programme was being implemented. This could have jeopardised the continuation of the programme – in particular the follow-up phases of the study. However, building trust with institutions separately through local collaborators proved to be very useful – it mitigated risks for the programme as institutions were satisfied with the team's responsiveness to their concerns and hence were still willing to participate.

Definitive stakeholders

Definitive stakeholders included MoES in Russia and AICTE in India. One of the recommended approaches to balancing the interests of definitive stakeholders is to integrate them in the decision-making or mediation process (Lampe, 2001). The programme team designed country-specific modules as part of such an approach. In India, for instance, the programme team received feedback both from policymakers and institutions that the quality of teaching in engineering colleges was a matter of concern for them. Based on this input, the programme team included a module in the academic staff survey that assessed the use of research-based teaching practices (such as active teaching) in STEM courses. Policymakers and institutions responded enthusiastically to the inclusion of this module and have expressed an ongoing interest in using it to improve pedagogy in engineering colleges. In Russia, the team has collected additional information on teacher qualifications as the MoES was interested in the impact of teacher qualifications on student outcomes.

Another example of this approach is the integration of the representatives of the stakeholder into the programme team in India. One of the incentives for policymakers in India to participate was to gain an understanding of how large-scale international assessments programmes are implemented. Specifically, they suggested building a 'unit'

within the AICTE that is devoted to continuing this programme of assessment and evaluation far into the future. Consequently, the programme team agreed to be a part of an ongoing effort to build capacity within the AICTE and collaborate with them to hire and train personnel for this unit. This has strengthened the commitment of Indian policymakers to engage with the programme.

The inclusion of definitive stakeholders into the decision-making process by allowing them to add topical survey modules or through the integration of their representatives into the project team had two main positive effects. First, it secured the support and participation of these stakeholders in the project by supporting their immediate goals and interests. Second, it helped to increase the level of policy relevance of the programme in each country.

Balancing the interests of multiple stakeholders across the decisions requires careful consideration how requests from some stakeholders can compromise trust or undermine the interest of other stakeholders. It requires communication and discussion of these concerns among stakeholders. Reporting of the results could be a good example of this issue. In Russia, the MoES was interested in getting the report with individual scores of universities in order to understand which institutions scored lower than others. This, however, could undermine the interest of the universities to get the objective results on how students learn. Similarly, universities were interested in getting the results of individual students but that would raise concerns among students on how these results will be used. That is why the research team agreed with the MoES and universities that it will provide the reports but they will not include the names of the universities (in the report for MoES) and the names of students (in the report for universities) as this will help to get more objective results and will serve their common interest.

Conclusion

The concept of quality assurance and its elements is constantly debated across nations. The need for more outcome-based elements in the quality assurance has created demand for large-scale assessments that measure a broad range of student skills especially in the context of decision-making at the international, national, and local (college and university) level. Despite this demand, stakeholders tend to be cautious about participating in large-scale assessment programmes because they are unclear whether the potential benefits will outweigh the perceived risks. This is especially true in an international context.

The fact that international assessments are vested in the interests of multiple stakeholders brings important trade-offs. First, universities and students participate in these assessments only if they are sure that their individual results won't be available to the public or other stakeholders. Large-scale assessments need to establish the right balance between restricting data dissemination and the actionability of their findings for higher education research and policy. The approach taken by the SUPER programme could inform other assessment initiatives. Second, international assessments will lead to policy changes only if there is at least some level of engagement by the decision-making bodies (usually ministries, funding or accreditation agencies) in the design and implementation of the assessment. However, these influential bodies can use their disproportionate power to attach rewards and punishments to the results of the assessments and thus undermine the interest of universities or students to

participate. The experience of the SUPER programme was that acquiring stakeholder buy-in is contingent upon four aspects: mitigating perceived risks, building trust, providing incentives for stakeholders to participate, and integration in the decision-making process. The programme team utilised a range of strategies to balance these aspects.

Eventually, most stakeholders agreed to participate because they believed that the benefits of the programme outweighed its potential risks. Only one college in each country declined to participate in the programme. Despite participation in the programme being voluntary, the response rates among students were also extremely high – 95% for China and India and 87% for Russia. Given these successes, we hope that the lessons from the SUPER programme in engaging stakeholders will be useful for policymakers and researchers seeking to conduct large-scale international assessments in the future.

Acknowledgments

Support from the Basic Research Programme of the National Research University Higher School of Economics and the China Scholarship Council is gratefully acknowledged.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the China Scholarship Council [201708410012]; Basic Research Programme of the National Research University Higher School of Economics.

ORCID

Guirong Li  <http://orcid.org/0000-0002-4745-8292>

Irina Shcheglova  <http://orcid.org/0000-0001-5949-9617>

Fei Guo  <http://orcid.org/0000-0002-0934-696X>

References

- Accreditation Board for Engineering and Technology. (2015). Criteria for accrediting engineering programs.
- Alexander, F.K. (2000). The changing face of accountability: Monitoring and assessing institutional performance in higher education. *Journal of Higher Education*, 71(4), 411–431.
- Altbach, P.G., Reisberg, L., & Rumbley, L.E. (2009). *Trends in global higher education: Tracking an academic revolution*. Paris: UNESCO Publications.
- Ashwin, P. (2015). Missionary zeal: Some problems with the rhetoric, vision and approach of the AHELO project. *European Journal of Higher Education*, 5(4), 437–444. doi:10.1080/21568235.2015.1087869
- Benneworth, P., & Jongbloed, B.W. (2010). Who matters to universities? A stakeholder perspective on humanities, arts and social sciences valorisation. *Higher Education*, 59(5), 567–588. doi:10.1007/s10734-009-9265-2

- Brahimi, T., Sarirete, A., & Ibrahim, R.M. (2016). the impact of accreditation on student learning outcomes. *International Journal of Knowledge Society Research (IJKSR)*, 7(4), 51–62. doi:10.4018/IJKSR.2016100105
- Carnoy, M., Loyalka, P., Dobryakova, M., Dossani, R., Froumin, I., Kuhns, K., & Wang, R. (2013). *University expansion in a changing global economy: Triumph of the BRICs?* Stanford: Stanford University Press.
- Donaldson, T., & Preston, L.E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review*, 20(1), 65–91. doi:10.5465/amr.1995.9503271992
- Eaton, J.S. (2003). Is accreditation accountable? The continuing conversation between accreditation and the federal government. CHEA monograph series, 1. *Council for Higher Education Accreditation*.
- Ellwell, P.T. (2012). A world of assessment: OECD's AHELO initiative. *Change: The Magazine of Higher Learning*, 44(5), 35–42. doi:10.1080/00091383.2012.706515
- Freeman, R.E. (1984). *Strategic management: A stakeholder approach*. Boston: Pitman.
- Froumin, I., Kouzminov, Y., & Semyonov, D. (2014). Institutional diversity in Russian higher education: Revolutions and evolution. *European Journal of Higher Education*, 4(3), 209–234. doi:10.1080/21568235.2014.916532
- Gu, L. et al. (2018). Validating the use of translated and adapted heighten® quantitative literacy test in Russia. In Eggins, H. (ed.), *Assessment of learning outcomes in higher education. cross-national comparisons and perspectives* (pp. 267–284). Cham, Switzerland: Springer.
- Jiang, K. (2015). Undergraduate teaching evaluation in China: Progress and debate. *International Higher Education*, 58, 15.
- Jongbloed, B., Enders, J., & Salerno, C. (2008). Higher education and its communities: Interconnections, interdependencies and a research agenda. *Higher Education*, 56(3), 303–324. doi:10.1007/s10734-008-9128-2
- Kapur, D., & Perry, E.J. (2015). *Higher education reform in China and India: The role of the state*. Cambridge: Harvard-Yenching Institute Working Paper Series.
- Kardanova, E., Loyalka, P., Chirikov, I., Liu, L., Li, G., Wang, H., ... Johnson, N. (2016). Developing instruments to assess and compare the quality of engineering education: The case of China and Russia. *Assessment & Evaluation in Higher Education*, 41(5), 770–786. doi:10.1080/02602938.2016.1162278
- Lampe, M. (2001). Mediation as an ethical adjunct of stakeholder theory. *Journal of Business Ethics*, 31(2), 165–173. doi:10.1023/A:1010749313936
- Leisyte, L., & Westerheijden, D.F. (2014). Stakeholders and quality assurance in higher education. In Eggins, H. (Ed.), *Drivers and barriers to achieving quality in higher education* (pp. 83–97). Rotterdam: Sense Publishers.
- Lennon, M.C., Fukahori, S., & Edwards, D. (2018). Global assessments of disciplinary learning outcomes: What we learned from AHELO. In Hazelkorn, E., Coates, H., McCormick, A. C. (Eds.), *Research Handbook on Quality, Performance and Accountability in Higher Education*, 264–274. Cheltenham, UK: Edward Elgar Publishing.
- Liu, O.L., Shaw, A., Gu, L., Li, G., Hu, S., Yu, N., ... Loyalka, P. (2018). Assessing college critical thinking: Preliminary results from the Chinese HEIghten® Critical Thinking assessment. *Higher Education Research and Development*, 37(5), 999–1014. doi:10.1080/07294360.2018.1467381
- Loyalka, P., Liu, O. L., Li, G., Chirikov, I., Kardanova, E., Gu, L., Ling, L., Yu, N. Guo, F. Ma, L., Hu, S., Johnson, A. S., Bhuradia, A., Khanna, S., Froumin, I., Shi, J., Choudhury, P. K., Beteille, T., Marmolejo, F. & Tognatta, N. (2019). Computer science skills across China, India, Russia, and the United States. *Proceedings of the National Academy of Sciences*, 116(14), 6732–6736.
- Lyytinen, A., Kohtamäki, V., Kivistö, J., Pekkola, E., & Hölttä, S. (2017). Scenarios of quality assurance of stakeholder relationships in finnish higher education institutions. *Quality in Higher Education*, 23(1), 35–49.
- Marmolejo, F., & Beteille, T. (2017). India: Is a college degree worth it? The world bank blogs. Retrieved from <http://blogs.worldbank.org/education/india-college-degree-worth-it>

- Mitchell, R.K., Agle, B.R., & Wood, D.J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, 22(4), 853–886. doi:[10.5465/amr.1997.9711022105](https://doi.org/10.5465/amr.1997.9711022105)
- Reynolds, S.J., Schultz, F.C., & Hekman, D.R. (2006). Stakeholder theory and managerial decision-making: Constraints and implications of balancing stakeholder interests. *Journal of Business Ethics*, 64(3), 285–301. doi:[10.1007/s10551-005-5493-2](https://doi.org/10.1007/s10551-005-5493-2)
- Shahjahan, R.A., & Torres, L.E. (2013). A ‘global eye’ for teaching and learning in higher education: A critical policy analysis of the OECD’s AHELO study. *Policy Futures in Education*, 11(5), 606–620. doi:[10.2304/pfie.2013.11.5.606](https://doi.org/10.2304/pfie.2013.11.5.606)
- Shaw, A., Liu, O.L., Gu, L., Kardonova, E., Chirikov, I., Li, G., ... Loyalka, P. (2019). Thinking critically about critical thinking: Validating the Russian HEIghTen® critical thinking assessment. *Studies in Higher Education*, 1–16, (in press). doi:[10.1080/03075079.2019.1672640](https://doi.org/10.1080/03075079.2019.1672640)
- Steedle, J.T. (2012). Selecting value-added models for postsecondary institutional assessment. *Assessment & Evaluation in Higher Education*, 37(6), 637–652. doi:[10.1080/02602938.2011.560720](https://doi.org/10.1080/02602938.2011.560720)
- Tilak, J.B. (2014). Private higher education in India. *Economic and Political Weekly*, 49(40), 32–38.
- Tremblay, K. (2013). OECD assessment of higher education learning outcomes (AHELO): Rationale, challenges and initial insights from the feasibility study. In Blömeke, S., Zlatkin-Troitschanskaia, O., Kuhn, C. & Fege, J. (Eds), *Modeling and measuring competencies in higher education* (pp. 113–126). Rotterdam: Sense Publishers.
- Tremblay, K., Lalancette, D., & Roseveare, D. (2012). *Assessment of higher education learning outcomes feasibility study report: Volume 1 – design and implementation*. Paris: OECD.
- Varghese, N.V., & Malik, G. (Eds.). (2015). *India higher education report 2015*. New York, NY: Routledge.
- Wang, L. (2014). Internationalization with chinese characteristics: the changing discourse of internationalization in China. *Chinese Education & Society*, 47(1), 7–26.
- Woodhouse, D. (1998). Quality assurance in higher education: The next 25 years. *Quality in Higher Education*, 4(3), 257–273. doi:[10.1080/1353832980040306](https://doi.org/10.1080/1353832980040306)