Education and Lifelong Development Research

Vol. 3, No. 1, p. 1-12, 2026

Original article

Improving program operations through managers' voices: Lessons from a national learning assessment initiative

Xiaoqing Wang[1]*, Chao Huang[2], Chenxi Song[3], Yurui Wang[2], Chengcheng Feng[1]

[1]*School of Education, Huazhong University of Science and Technology, Wuhan 430074, China*
[2]*The School of international education, Huazhong University of Science and Technology, Wuhan 430074, China*
[3]*Funing County Bureau of Industry and Information Technology, Yancheng 224400, China*

Keywords:
College student learning assessment
assessment program
program manager
Chinese college students

Cited as:
Wang, X. Q., Huang, C., Song, C. X., Wang, Y. R., & Feng, C. C. (2026). Improving program operations through managers' voices: Lessons from a national learning assessment initiative. Education and Lifelong Development Research, 3(1): 1-12.
https://doi.org/10.46690/elder.2026.01.01

Abstract:
Currently, over 200 institutions participate annually in China's five major college student learning (CSL) assessment programs every year; however, there remains a lack of systematic research focusing on the operation of these initiatives. This study examines the Mingde Program—one of the longest-running and largest-scale national CSL assessment programs—as a representative case. A mixed-methods approach was employed, combining questionnaire surveys with semi-structured interviews. Questionnaires were distributed to program managers from 43 higher education institutions, and follow-up interviews were conducted with 25 program managers. The findings reveal several key challenges in the program's implementation including limited dissemination and use of assessment results, insufficient professional capacity and turnover among program managers, and inadequate support services provided by the research team. Nonetheless, this program has demonstrated a positive impact on institutional development. Based on these findings, five recommendations are proposed: (1) broaden the dissemination of program results; (2) enhance utilization of assessment outcomes; (3) strengthen the professionalization of program managers; (4) ensure the stability of program management personnel; (5) develop responsive professional services according to institutional needs.

1. Introduction

In the 1990s, American state legislatures pressured universities to take more responsibility for college student learning (CSL) amid declining graduation rates, rising education costs, low economic returns and high public concern (Klein et al., 2005). The federal government also made student learning a primary task for certification agencies, leading 44 states to create accountability systems, most referencing CSL (Naughton et al., 2003). Consequently, undergraduate learning assessment program with academic research gained prominence (Astin & Antonio, 2012). Programs like the National Survey for Student Engagement (NSSE), College Student Experience Questionnaire (CSEQ), the Student Experience in the Research University (SERU) and other assessment programs later developed rapidly and spread globally (Coates & McCormick, 2014; Yin & Zheng, 2017).

China was one beneficiary. Since 2000, five major assessment programs have developed promptly: Peking University's National University Teaching Quality and Student Development Monitoring Program ("Peking University Program" for short), Tsinghua University's China College Student Survey (CCSS), Xiamen University's National College Student Survey (NCSS), Huazhong University of Science and Technology's Student Survey of Learning and Development (SSLD) and SERU Program. Annually, over 200 colleges and universities participate, about one-tenth of Chinese institutions (Wang & Yin, 2022). Furthermore, the indicator system of the Implementation Plan for the Audit and Evaluation of Undergraduate Education and Teaching in Ordinary Institutions of Higher Learning (2021-2025), focuses on students' satisfaction and growth (Ministry of Education of the People's Republic of China, 2021), ensuring continued rapid growth in program

Yandy Scientific Press

*Corresponding author.
Xiaoqing Wang and Chao Huang have contributed to the workequally and should be regarded as co-first authors.
E-mail address: wangxiaoqing@hust.edu.cn (X. Wang); huangchao2005@hust.edu.cn (C. Huang).

Received October 11, 2025; revised December 25, 2025; accepted January 15, 2026; available online March 20, 2026.

participation.

After years of operation in Mainland China, how are these programs running in the participating colleges and universities? What positive effects and problems exist? This research addresses the above questions by employing a mixed research method, using the Mingde Program (a pseudonym), as a case study. The Mingde Program was selected for its longevity and scale from among several prominent large-scale student survey initiatives. It examines teachers' teaching and students' development. The assessment object is for college students. Originally a research program in Xin Province (a pseudonym), it is now overseen by local education authorities with long-term funding. Evolving from explicit funding to implicit policy support, it has operated for over a decade, becoming China's longest-running such program, engaging about 70 institutions annually, and transitioning to a national program since 2017[1].

Commissioned by the Mingde Program team, this study research investigates its operational status within participating institutions, analyzes existing problems, and proposes countermeasures. The findings aim to help institutions better utilize program results to improve the quality of undergraduate education (QUE), enable the Mingde Program to provide more precise academic services, thereby contributing to the assurance of QUE in Chinese higher education.

2. Literature review

Established student experience assessment frameworks, such as NSSE and CSEQ in the United States, the UK's National Student Survey (NSS), and Australia's Course Experience Questionnaire (CEQ), demonstrate that international assessment programs contribute to institutional development through internal influences and external applications (Center for the Study of the Student Experience, 2019; Griffin et al., 2003; National Survey of Student Engagement, 2021; The Student Survey, 2026). Internally, these programs promote teacher reflection on teaching practices and aid students in self-positioning. Externally, the results are used in institutional certification systems and inform college choices for students and parents.

Systematic research on the impact of programs like NSSE has been conducted by scholars such as Han (2012) and Kuh (2020), while Zhang (2014), along with Lv & Chen (2010) and Bao (2014), has contributed to understanding the internal and external influences of the CSEQ, NSS, and CEQ. Regarding internal influence, the application in college teaching is relatively direct, though impacts on teachers and students vary. In the United States, NSSE and CSEQ have a better impact on students than that of the programs in Commonwealth countries, with results assisting students in reflecting on their learning. In contrast, a distinctive feature of the Australian CEQ is that its results directly inform and promote teaching reflection among instructors.

Concerning external applications, two American programs are often integrated into institution certification systems, significantly enhancing the scientific rigor of the teaching administration's intervention in the caliber of collegiate instruction. Conversely, the British government follows the "limited use principle" for the NSS results, avoiding direct intervention in institutional teaching. Meanwhile, the Australian program's results are directly tied to the government rankings of university teaching quality and financing allocations. Furthermore, most program outcomes provide a foundation for students and parents when selecting a college. Similar programs in Canada (Norrie & Conway, 2014), Ireland (Drennan et al., 2014), and South Africa (Strydom & Mentz, 2014) have also positively contributed to the QUE assurance system.

In contrast, systematic studies on the impact of domestic programs within China are relatively scarce. There are three categories of China's institutions participating in the CSL programs. The first category involves participation in international programs, such as SERU program of the University of the UC- Berkeley (Chang, 2022). This has attracted institutions like Nanjing University, Xi'an Jiaotong University, Hunan University, Tongji University, Peking University and Dalian University of Technology. The second category encompasses involvement in localized programs, such as the Peking University program, CCSS, NCSS, and SSLD. The third category consists of independently developed school-based programs, such as the baseline survey of freshmen at Peking University (Zhu & Zhang, 2020), and the survey of Undergraduates' Learning Situation in Sun Yat-Sen University (Qu et al., 2013).

Research on these programs is emerging. Regarding internationalization programs, Wang & Wang (2018) have demonstrated that Nanjing University utilized the SERU program as one of the troikas of teaching reform, sparking empirical studies and fostering an evidence-based culture. For localized projects, Wei & Chen (2016) discovered that the SSLD program to be a useful instrument for understanding the impact mechanism of undergraduates learning effects. Concerning the CCSS, Shi & Wang (2018) noted that it provides an empirical basis for self-assessment and university assessment of students' learning process and gains. A more prominent role is reflected in its integration into Tsinghua quality management system (Shi, 2016). Published studies on the impact of school-based programs, however, remain lacking.

In summary, while long-term investigations and relatively rich findings exist on the effects of the CSL assessment programs both domestically and abroad, research methods remain concentrated on conventional "text-to-text" paradigms. Field research, particularly comprehensive and systematic studies combining quantitative and qualitative data collection, is still scarce. Furthermore, previous studies have predominantly focused on the positive benefits of these programs, paying insufficient attention to the operational challenges. To address these gaps, this study uses the Mingde Program—a large-scale, long-running Chinese initiative—as a case study. Employing mixed methods research, it examines the program's current operational state, identifies existing challenges, and proposes targeted remedies.

3. Research design

3.1 Theoretical framework

This study aims to understand the effects of the Mingde Program within participating institutions, an inquiry that falls

under meta-evaluation. While a full meta-evaluation is not feasible under the current research constraints, the framework for the questionnaire is guided by established meta-evaluation standards. Specifically, it draws upon frequently asked "U.S. meta-evaluation questions" proposed by the Joint Committee on Standards for Educational Evaluation (JCSEE), which concern the clarity of findings, sufficiency of data, effective dissemination, and application of results (Sanders, 2012). For instance, "Are the program evaluation's findings clear? Are the data and information sufficient to justify the assessment conclusions? Have the evaluation results been shared effectively? Have the outcomes been applied? etc."

The foundational JCSEE meta-evaluation standards—Utility, Feasibility, Propriety, and Accuracy—have been widely adopted and adapted internationally, including in China (Yan, 2010; Zhang, 2014). The third edition of the JCSEE standards further introduced a fifth dimension: Evaluation Accountability (Joint Committee on Standards for Educational Evaluation, 2019). Among these, the Utility standard aligns most closely with this study's focus on program impact. This standard encompasses critical aspects such as dependability of the evaluators, stakeholder identification, negotiation of evaluation goals, value judgments, information screening for stakeholders, meaningful evaluation processes and results, timely and accurate report communication, evaluation consequences, and influence issues.

Consequently, the dimensions of this study are structured around seven key aspects derived from this utility-oriented framework (See Table 1). These dimensions directly inform the design of the questionnaire and the interview protocol for the subsequent mixed-methods investigation.

3.2 Data collection and analysis

The most popular consistency parallel design for mixed method research was utilized in this study, which collected both quantitative and qualitative data concurrently and combined the findings of the two analyses to make the findings more convincing (Creswell & Clark, 2017). According to program archives, the Mingde program involves an average of 70 national universities annually, with approximately 50 located within its originating Xin Province. Methodologically, the research employs a mixed-methods approach in which a questionnaire survey serves as the primary quantitative component, supplemented by qualitative insights obtained through in-depth interviews.

3.2.1 Quantitative part

The quantitative data collection was conducted at the end of May 2018 during two separate program promotion meetings. One meeting targeted "old member" institutions from Xin Province, while the other targeted "new member" institutions from outside the province participating for the first time or for a year. The survey unit was the participating college or university. Questionnaires were distributed to all university-level program leaders or attendees of the kick-off meetings, with one respondent per institution. These individuals are collectively referred to as "program managers" in this research.

A full sample survey was administered. Sixty-six questionnaires were distributed, to the participants of the program kick-off meetings, and 53 were returned (80.3% recovery), and 45 were deemed valid (84.9% effective rate). After removing two polytechnic college, the final quantitative sample consisted of 43 undergraduate institutions. This sample included 25 institutions from within Xin Province (58.1%) and 18 from outside the province (41.9%). Program 985 and Program 211 institutions constituted 57.5% of the sample, while local undergraduate institutions accounted for 42.5%.

The individual features of program managers were as follows: 70% were from student affairs departments, 10% from teaching affairs, 7.5% from higher Education research institutes, 4.7% from teaching quality assessment offices, and 7.8% from other departments. Gender distribution was 59.5% female and 40.5% male. In terms of administrative rank, 62.5% were section-level cadres or above, and 37.5% were section staff and others. Regarding professional titles, 36% held senior titles, 40% intermediate, and 24% junior or other. Most managers (79.3%) had been responsible for program for two years or less, 13.8% for three years, and 6.9% for over five years.

In terms of reliability, Cronbach's alpha coefficient was used to test the program effect (0.696), degree of importance to the program by the institutions (0.599), dissemination scope of program results (0.540) and other dimensions. Generally, the α coefficient exceeded 0.6. It is currently allowed for it to be more than 0.5 because there are only two precise indicators of the degree of importance to the program by the institutions and dissemination scope of program results (SPSSAU, 2020).

In terms of validity, this study tested content validity, and criterion-related validity. First, the content validity was established through alignment with the Utility standards of worldwide meta-evaluation from which the questionnaire's structure was derived. Second, criterion-related validity was assessed by examining concurrent validity. The project manager's understanding of the program was used the criterion variable, and its correlation with the university's emphasis, the satisfaction of the department leader, the satisfaction of the responsible university leader, and program impact was analyzed. The results revealed a statistically moderate correlation ($p<0.01$) between these four indicators and the criterion variable, demonstrating adequate criterion-related validity of the questionnaire.

As mentioned above, the annual average number of institutions participating in Mingde Program in China is about 70. Given that the survey sampled one manager from each institution participating in the program, the resulting sample size is objectively small. This determines that the quantitative analysis primarily focuses on descriptive statistics. Chi-square tests were employed to examine significant differences in key indicators based on institutional location (in-province vs. out-of-province), and tier (Program 985/211 vs. local undergraduate colleges), providing a general understanding of the program's differential impact.

Table 1. The framework for the questionnaire and the interview protocol.

Dimension	Sources
1) Attitude (emphasis and project cognition)	Reflection of researchers
2) Purpose of participation in evaluation	Sanders (2012)'s "The program Evaluation Standards", The meta evaluation standards of German Evaluation Association and Swiss Evaluation Association
3) Information matching degree/ question feedback	"U.S. meta-evaluation questions"
4) Feedback evaluation report (report clarity, explanation and interaction with experts)	"U.S. meta-evaluation questions", Sanders (2012)'s "The program Evaluation Standards"
5) Result dissemination (mode and scope)	Literature review, Reflection of researchers, expert opinion
6) Application of results (application, etc.)	Literature review, "U.S. meta-evaluation questions", The meta evaluation standard of German Evaluation Association
7) Effect (changes caused by follow-up measures, effect, satisfaction, whether to continue to participate, whether to recommend to others)	Evaluation Standards for Educational Programs (1994): U7 Evaluation Impact, Reflection of researchers

Source: Zhang (2014): "Teaching quality evaluation and guarantee based on the value-added development of students", pp. 185-187; Sanders (2012): "The program evaluation standards - How to assess evaluation of education program" (Second Edition), pp. 61-62.

3.2.2 Qualitative part

To enhance the explanatory power of the findings, qualitative interviews were conducted with 25 program managers from 15 public institutions (coded A1-A15) among the 43 institutions between May 2017 and January 2019. The interviewees were selected from 5 Program 985 institutions, 5 Program 211 institutions, and 5 local undergraduate institutions, considering a balance of arts, science and engineering, and comprehensive expertise. A total of 25 teachers were interviewed, including 7 males and 18 females. Sixteen were program leaders, with the rest being participants and department/ college leaders. Interviews were primarily face-to-face, averaging 40 minutes each.

The sample coding is based on the combination of the respondent's function and the university. For example, "A1M1", where "A1" stands for the university, and "M1" stands for the program leader; M0 and M1 respectively refer to the former and the current person in charge. "A1P1" indicates that the teacher is a program participant of A1 University, "AH1" a head of Mingde Program team. "AS" a program supervisor, and "AT1" an expert familiar with the program. The host institution is referred to as "Mingde University". Interviewees also included 16 research team leaders, core participants, and experts from Mingde Program, Qinmin Program (pseudonym) and Zhishan Program (pseudonym). The three programs presented here—Mingde Program (using code 'A' for universities), Qinmin Program (using code 'B'), and Zhishan Program (using code 'C')—are anonymous processing for three of the above mentioned five Chinese CSL assessment programs.

To complement the quantitative component, the interview data were analyzed using an inductive analytical approach. Drawing on principles of open coding (Strauss & Corbin, 1990), key themes were identified from interview transcripts. Concurrently, a deductive analysis was employed to examine themes in alignment with the utility-oriented framework in Table 1.

3.3 Research ethics

Ethics considerations were strictly adhered to. Anonymity was guaranteed for all survey and interview participants. In qualitative research, ethical concerns were reflected in that researchers needed to seek the consent of respondents in advance and keep the information provided by them strictly confidential (Chen, 2000). We respected interviewees' preferences regarding audio recording and engaged in member checking by discussing findings with several respondents to ensure accurate interpretation and application of the data.

4. Research findings

This section presents descriptive statistics across eight aspects: attitude towards the program, reasons for participation, feedback about the university, result dissemination, result use, program effectiveness, program feedback, and others. which including in three categories: overall, universities in the province, and universities outside the province. It should be emphasized that as new program managers in some universities took office, they did not know much about the operation and function of the project, so most of the questions were set as "do not know".

4.1 Attitude towards the program

Institutional attitude was measured through perceived importance and managerial understanding. Nearly 80% of colleges and universities attach great importance to participating in Mingde Program, with the remainder considering it "general." In-province institutions valued the program more highly, likely due to administrative mandates, though a chi-square test showed no statistically significant difference in importance between the two groups. Regarding program understanding,

Table 2. Chi-square analysis of program understanding by manager location (In-Province vs. Out-of-Province)

	Value	df	Progressive Sig. (Bilateral)	Exact Sig. (Bilateral)	Exact Sig. (Unilateral)
Pearson Chi-square	2.978[a]	1	0.084*		
Continuous correction [b]	2.007	1	0.157		
Likelihood Ratio	3.022	1	0.082		
Fisher's exact test				0.124	0.078
Linear and linear combination	2.909	1	0.088		
N in valid case	43				

Note: ∗ means significant at the 0.1 level.

nearly 50% of project managers had a "good" grasp, over 40% "general", and nearly 10% "do not know." Understanding was higher among in-province institutions, which mostly surpassed the "pass line," compared to just over 30% of out-of-province managers reporting "good" understanding. As is shown in Table 2, a significant difference in understanding existed between the groups ($p < 0.1$), though the effect size was modest (Cramer's V = 0.263) (Muijs, 2004). This disparity is primarily attributed to the longer participation history (over ten years) of in-province institutions, unlike the one-year maximum for out-of-province members.

4.2 Reasons for participation

According to the survey (See Fig. 1), the primary reasons for participation: first, nearly 90% of institutions cited the program's utility in providing feedback on teaching quality and student development; second, nearly half acknowledged the research strength of the program team; third, participation was required by superior departments or was free of charge.

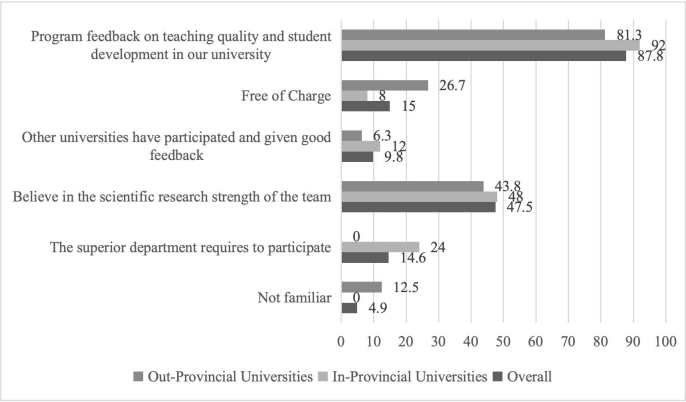

Fig. 1. Distribution of reasons for institutions to participate in Mingde Program (%, N=16, 25, 41).

The first reason reflects a genuine need, the second indicates recognition of the program's academic credibility, and the third was specific to in-province institutions, which were required by local educational authorities to participate. The perspective of interviewee A3M1 supports these findings:

"I see benefits in this program's standardization, rollability, and long-term nature. It is entrusted by Xin Province education authority to Mingde University. On the one hand, there is administrative participation; on the other hand, it's delegated to the School of Education of Mingde University, a high-quality, high-level educational research unit, to get a double guarantee. From the project quality itself, it should be very good, very high. In terms of the development and construction of our school itself, we also need such a data monitoring project, and our school also attaches great importance to it."

For in-province institutions, administrative requirements became a secondary motivator as the program met their internal needs, suggesting a potential shift from external pressure to internal rational adoption. The "free of charge" option was primarily selected by out-of-province institutions, likely related to the higher proportion of local undergraduate institutions with tighter budgets. Cross-analyzing of "whether it is an institution in the province", "type of institution", and "free of charge" have demonstrated this argument. The five institutions selecting "free" were all local undergraduate colleges. Unlike the free Mingde Program, others like the Qinmin Program, the NSSE and Zhishan Program adopt a market model with fees (C2M1), a shift that impacted its institutional scale after transitioning from free to paid (BH1).

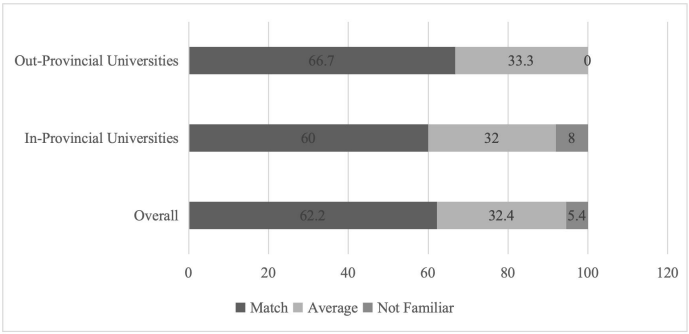

Fig. 2. The matching degree between the information feedback by Mingde program and the actual situation of the institutions (%, N=12, 25, 37).

4.3 University situation feedback

Regarding information matching—a prerequisite for effectively disseminating and utilizing program results—the survey (See Fig. 2) indicates that nearly 60% of all institutions consid-

Table 3. List of Problems in Mingde Program Alliance Institutions.

Concern Dimensions	First Place (%)	Second Place (%)	Third Place (%)
1. Main problems of institutions (N=36)	CL (55.6)	AA (33.3) CP (33.3)	CTE (27.8)
2. Dimensions of program concern			
a. CL(N=36)	Learning motivation(55.6)	Time allocation in/out of class (47.2)	Professional identity (38.9)
b. EL (N=37)	Participation in extracurricular activities (70.3)	University adaptation (54.1)	Training of student cadres (10.8)
c. CTE (N=37)	Students' satisfaction with teaching quality (43.2)	Communication between teachers and students after class (40.5)	Students' preference for teaching methods (37.8)
d. SW (N=36)	Student support (e.g., orientation guidance) (54.1)	Communication between counselors and students (50)	Students' satisfaction with counselors (29.7)
e. AA (N=36)	Students' General abilities (63.9)	Students' professional abilities (41.7)	Students' moral values (27.8)
f. CP (N=36)	Individual development planning (61.1)	Graduation plan and destination (30.6)	Student employment level (27.8)

Notes: CL= course learning; EL= extracurricular life, CTE= course teaching evaluation, SW= student work, AA= academic achievements; CP= career planning. The sample size for these dimensions ranges from 36 to 37, reflecting minor variations attributable to missing responses for specific items.

ered the feedback from Mingde Program accurately reflected the actual situation. Out-of-province institutions expressed slightly stronger agreement (approximately 70%) compared to in-province institutions (about 63%). However, chi-square analysis confirmed no statistically significant difference between the two groups.

In terms of problem feedback, the program primarily focuses on students' course learning (CL), extracurricular life (EL), course teaching evaluation (CTE), student work (SW), academic achievements (AA), and career planning (CP). It has been found that the top three problems of institutions (See Table 3). Institutions mainly struggled with issues related to CL, AA, CP, and CTE. This consistency underscores the relevance of the program's focus and highlights the significance of CSL assessment.

Overall, chi-square tests revealed no significant difference between in-province and out-of-province institutions concerning broader academic learning issues. When comparing specific indicators, a notable exception was the communication between counselors and students (one indicator of SW), where a slight but significant difference emerged (χ^2 = 3.273, df=1, N=36, p = 0.07, φ = -0.302), with the issue being more pronounced in out-of-province institutions. Nevertheless, the weak effect size suggests this difference is minimal. In fact, the national and regional reports in 2017 showed that regional performance in student-counselor interaction was comparable to or even slightly better than the national average on key indicators, such as counselors aiding moral/personal development and students proactively seeking advice[2].

In essence, whether examining major student learning challenges or specific dimensional, the problems identified are largely consistent between institutions inside and outside the province. Based on the national and regional reports in 2017, it can be concluded that these issues appear universal across the national context. For instance, over 40% of students had a negative attitude toward consulting teachers or counselors about academic or personal matters. Furthermore, the data revealed ethically concerning trends, including significant proportions of students endorsing individualism (36.0%) or prioritizing material gain above all (20.9%), indicating the presence of morally unbalanced student subgroups[2].

4.4 Dissemination of results

The dissemination of Mingde Program's findings is examined through two dimensions: mode and scope. Regarding dissemination mode, the survey indicates that 70% of institutions shared results via paper documents, while approximately 30% utilized online channels. The distribution between in-province and out-of-province institutions was similar, with chi-squared tests confirming no significant difference between the two groups. While paper-based dissemination ensures clarity within a limited scope, online sharing holds the potential to broaden the reach of results.

In terms of dissemination scope, as illustrated in Figure 3, the feedback reports were primarily circulated internally. Overall, nearly 80% of responsible departments submitted the report to university leaders; 80% studied within their own department; 60% sent it to the faculty. However, only 20% shared the report with other relevant departments, and very few institutions made reports accessible to all teaching staff via platforms like office automation (OA) systems. Notably, no institution provided results to all students, staff, or external audiences. Chi-squared test confirmed no statistically significant difference in dissemination scope between in-province and out-of-province institutions, with both concentrating distribution among senior administrators, their own departments

and faculties. This is in line with the presumption that the program findings are only distributed locally.

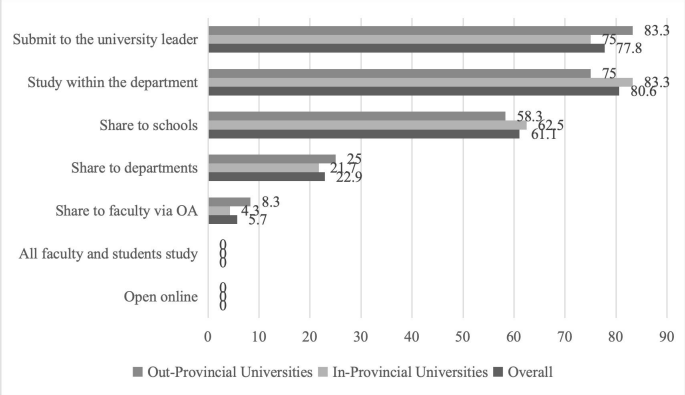

Fig. 3. The dissemination scope of Mingde program feedback report in institutions (%, N=12, 24, 36)

Beyond quantitative scope, the question of whether results are shared in full or selectively remains. Interviews with program managers revealed that disseminated content typically consisted of processed data—often benchmarked against the institution's historical data or provincial/national norms (A3M1). Perspectives varied on the completeness of shared content: some claimed that universities selectively highlighted positive outcomes when reporting to leaders, key meetings, or other departments (AH1), while others maintained that their feedback results included both achievements and identified problems (A3M1).

Application of results ideally, the institutions would utilize Mingde Program feedback to build on successes and address identified problems, thereby promoting teaching and learning. However, a gap persists between this ideal and actual practice. Regarding methods of results use, the survey (See Fig. 4) shows that nearly 80% of institutions conducted internal departmental reviews of the findings. Over 40% undertook specialized research on prominent problems, while over 30% incorporated these issues into departmental annual plans. Nearly 20% integrated them into university -level annual plans, nearly 10% applied results in other ways, and only one institution reported no usage. Disaggregated data shows in-province institutions focused strongly on departmental review (nearly 90%), though fewer (only 30%) conducted specialized research or integrated issues into the departmental plans. In contrast, out-of-province institutions showed a low rate of internal review (under 60%) but higher engagement in specialized research (over 60%) and departmental planning integration (nearly 30%). This indicates both groups have room for improvement in solving prominent problems.

Qualitative data illustrates effective application where institutions conducted focused problem research, using program data to support university decision-making. For instance, one institution implemented targeted online loan guidance after data revealed 30% of students frequently shopped online (A12M1). Such cases demonstrate that major issues identified through the program can receive serious institutional attention and trigger concrete actions.

As an education and teaching reform too, the survey found 65.8% institutions operated their own internal research programs for monitoring teaching quality and student development. Despite this, nearly 30% reported using Mingde Program specifically to drive education reform. While this percentage may appear modest, it remains significant given the complex, and systematic nature of such reforms. Conversely, nearly 30% did not use the program for reform purposes, and over 40% of program leaders or participants were unaware of its application in this context. In-province institutions were 28 percentage points more likely to use the program as a reform tool than out-of-province institutions.

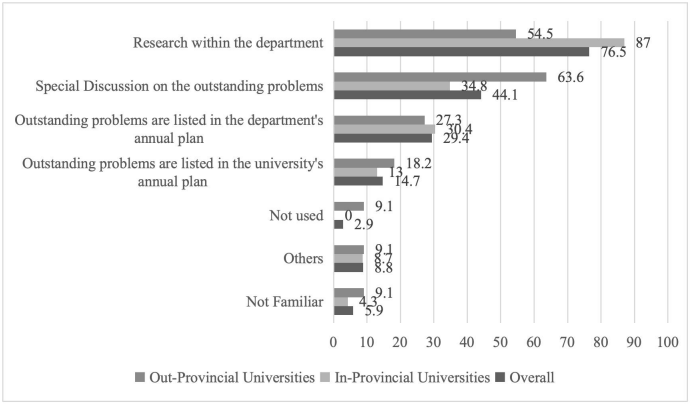

Fig. 4. Institutional use of Mingde program feedback report (N=11, 23, 34)

Interview data confirms the program's substantial impact on some universities' development. For instance, the research report supported A2 University's "Double-First-Class" initiative strategy (A2M1):

"When the 2017 report came out, we were conducting 'Double- First-Class' research, including an education quality research report. Each department had to think about how to construct it in terms of talent cultivation. In this process, we happened to use this data. The good thing about our data is that it compares the colleges and universities on specific themes."

Ultimately, regardless of the application level, the key lies in actively utilizing the results. As one manager emphasized, "Use is effect" (A15M1).

4.5 Program effectiveness

This study assessed the effectiveness of the Mingde program through its impact on education and teaching reform, satisfaction of department leaders and university leaders, sustained participation, and willingness to recommend it to peers.

Regarding the program's impact on education and teaching reform, one fifth of institutions considered it highly effective, over 30% reported a generally good effect, and only 10% perceived little effect. However, more than 30% of program managers were unaware of its impact (See Fig. 5). A comparative analysis shows that nearly 30% of in-province institutions viewed the program as highly effective, indicating stronger perceived benefits among this group.

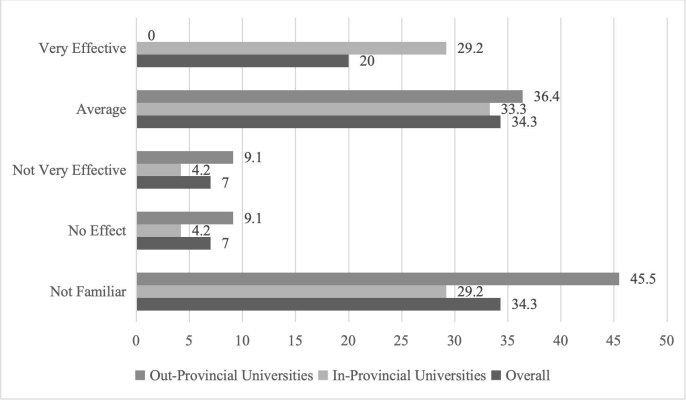

Fig. 5. The Impact of Mingde Program on the Educational and Teaching Reform of institutions (%, N=11, 24, 35)

In terms of satisfaction with program outcomes, more than half of the department leaders were satisfied, over 20% were moderately satisfied, and nearly one-third of managers were unaware of their department leaders' views. Satisfaction among department heads in in-province institutions was approximately 13 percentage points higher than in out-of-province institutions. Similarly, regarding the satisfaction of the school leaders, nearly half of all institutions reported relatively high satisfaction, and 20 percent were moderately satisfied. In-province institutions again showed higher satisfaction rates by nearly 14 percentage points. Notably, around 50% of managers in out-of-province institutions were unaware of the satisfaction of their university leaders' satisfaction, suggesting potential limitations in internal communication and application of program results within these institutions.

The program demonstrated strong engagement in terms of sustained participation and recommendation. Ninety percent of program managers confirmed their institutions would continue to participating, with nearly 90% of in-province and over 90% of out-of-province institutions expressing this intention. Similarly, over 90% of respondents would recommend the program to other universities, including 90% from in-province and 100% from out-of-province institutions. Chi-square test indicated no significant differences between the two types of institutions in these dimensions, reflecting consistently high levels of continued commitment and advocacy across the program alliance.

4.6 Program feedback

Feedback on the program's content indicates that 85% of managers believed that Mingde Program questionnaire reflects current educational priorities. This approval reached nearly 90% among in-province institutions, compared to slightly under 80% for out-of-province participants. Similarly, regarding questionnaire comprehensiveness, nearly 90% of all respondents considered the content sufficient, while over 10% suggested expansions. Proposed additions included indicators for institutional characteristics, learning methods (e.g., online, cross-cultural, and advanced learning), internship participation analysis, talent cultivation positioning, expanded teaching behavior metrics, campus cultural atmosphere, government/societal support, and adaptation challenges of ethnic minority students[3]. Chi-square tests showed no significant differences between in-province and out-of-province institutions across these content evaluation dimensions.

Regarding interaction with program experts, the survey revealed that 40% of institutions maintained moderate communication with the Mingde Program team, 20% interacted infrequently, and only about 10% engaged regularly. Overall, nearly 70% did not frequently interact with the program team. As one program supervisor (AS) noted, this may be because top-tier universities (Program 985 and 211) often possess strong internal research teams capable of independent data analysis, reducing their reliance on program support compared to local institutions. Notably, the program team has enhanced its service model in recent years by introducing customized support, which has deepened engagement with participating universities.

In terms of service suggestions, institutions proposed seven main improvements (See Fig. 6): a national annual report (72.5%), longitudinal data analysis (70%), professional lectures and research consulting (53.8%), benchmarking by institution type (50%), sharing of successful experiences (47.5%), expanded distinctive education indicators (42.5%), and deeper cooperation (35%). Some also requested training in research methodology.

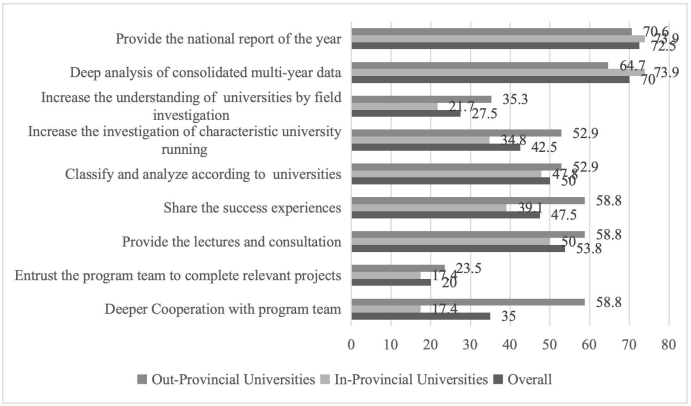

Fig. 6. Adding service proposals of Mingde program which institutions hope (%, N=13, 23, 43)

The Mingde Program team already provides several of these services, including national reports, institutional datasets, methodology training, and specialized support. Universities particularly value benchmarking capabilities—comparing their performance against peer and national norms. Previously, institutions relied on the team for provincial reference data, but recent efforts toward full electronation aim to enable independent, efficient cross-sectional and longitudinal analysis via an online platform (AS). The team also delivers professional lectures and consultation, such as training sessions for counselors at provincial training base (AH1).

Areas like experience sharing and distinctive education assessment represent future development directions. Such practices align with services provided by established organizations like the Educational Testing Service (ETS) (Bogue & Hall, 2008). While service requests were generally consis-

tent between in-province and out-of-province institutions, the latter expressed stronger need for "close interaction with the program team." A slight statistical difference was observed (7.376[a], df=1, N=40, p=0.007, φ=-0.429), though overall service recommendations showed no significant variations between the two groups.

Other Findings Beyond the core dimensions, this study compared the conditions responses across institutional tiers—Program 985/211 universities and local undergraduate colleges. Chi-square test revealed significant differences in only two service suggestions: "field investigation of the institution" (5.290a, df=1, N=39, p=0.021, φ=-0.368) and "in-depth cooperation with the project team" (3.804a, df=1, N=39, p=0.051, φ=-0.312). However, the strength of association was weak in both cases.

An examination of five program result indicators showed that program 985/211 universities generally demonstrated more pronounced effectiveness than local institutions, particularly regarding impact on education and teaching reform. Nevertheless, these differences did not reach statistical significance across the five indicators. Similarly, while elite universities held a modest advantage in the dissemination and application of program results, no significant differences emerged between the two institutional types, and detailed data are not presented here.

In summary, the study confirms Mingde Program's positive impact on institutional development. Key evidence includes the high importance universities attach to participation, their substantial program awareness, and the localization of evaluation practices enabling widespread internal dissemination of results. Notably, nearly one third of institutions were using this program as a tool for education reform. These findings collectively demonstrate the program's implicit impact on teaching and learning across participating institutions, with minimal variation across regions or university types.

Consequently, these findings challenge two prevailing assumptions: first, that the program exerts little to no institutional influence (AT2); second, that its impact varies significantly across institutional tiers, with local undergraduate colleges benefiting more substantially than elite universities (AS). Despite these positive outcomes, the study identifies persistent challenges requiring attention, particularly concerning the dissemination and application of results, the professional capacity and stability of program managers, and the need for enhanced program team services.

5. Countermeasures and recommendations

Based on the systematic investigation into the operational challenges facing the Mingde Program, this study proposes five strategic recommendations aimed at enhancing program effectiveness and sustainability. These evidence-based suggestions address critical areas requiring coordinated efforts between the program team and participating institutions, with particular attention to the cultural, administrative, and professional dimensions identified through the research.

5.1 Broadening the dissemination of program results

The study reveals significant limitations in current dissemination practices, particularly regarding inter-departmental sharing. This constraint stems from two primary factors: internal departmental competition (like between the student work department and the teaching services department) and the cultural concept of Chinese tradition that "domestic ugliness cannot be publicized". Given the program's comprehensive evaluation framework addressing both academic and extracurricular dimensions, restricted dissemination substantially barriers maximizing program utility.

Comparative data from international assessment programs underscores this gap. The NSSE maintained 75-80% sharing rates with faculty (Wang & Niu, 2022), while Mingde program achieves only 5.7%. Furthermore, as Astin's (1985) theory of educational value-added constitutes the theoretical foundation of the program, the ultimate beneficiaries—teachers and students—should access results to inform their practices. The Zhishan Program's experience illustrates how institutional apprehension about revealing shortcomings frequently constrains dissemination, diminishing potential impact on educational quality.

5.2 Enhancing utilization of program results

The investigation shows program results primarily serve limited purposes, focusing on departmental reviews rather than strategic planning. Since meaningful utilization justifies institutional participation (Ewell, 1985), this gap demands systematic attention. The program's findings are mostly used for departmental internal research, to examine pressing problems, and to a lesser extent, to incorporate into departmental or even university plans. However, one of the goals of institutions participating in Mingde program is to promote the development of the schools and students, which is consistent with the program's service goal[2].

This challenge mirrors international experience. The Wabash National Study documented only 25% institutional response rate (Blaich & Wise, 2011), while NSSE data similarly indicated underutilization (Wang & Niu, 2022). To bridge this gap, institutions should develop systematic mechanisms for translating findings into actionable strategies, ensuring alignment between objectives and usage patterns across administrative and academic units.

5.3 Strengthening professional capacity of program managers

Service requests predominantly relate to research literacy needs, particularly data analysis capabilities. Because of the lack of professional data statistics in these institutions, they hope that Mingde program team can provide more in-depth services. Given the program team's limited resources against unlimited institutional demands, building local capacity becomes essential for sustainable impact. The proverb "To teach others the way of fishing is far better than to present them fish" captures the needed strategic approach.

While maintaining existing research methodology training—including implementing annual workshops requested by managers (A3M1)—institutions must simultaneously establish sustainable professional development mechanisms. A competent management team can better interpret reports, conduct meaningful secondary analyses, and serve institutional decision-making needs more effectively. This dual approach addresses both immediate service gaps and long-term sustainability concerns.

5.4 Ensuring stability of program management personnel

High turnover among program managers poses significant challenges, with most serving less than three years. This mobility manifests concerningly: 8.9% of program managers did not know about the program; 29.7% and 24.3% respectively were unaware of the satisfaction of the university leaders and department leaders; 40.5% didn't know about the program's teaching reform role; and 24.3% were not sure about the frequency of their interaction with the expert group. Interviews (A2M1, A11M1) confirm frequent changes necessitate repeated reorientation and involve inadequate knowledge transfer.

This instability undermines dissemination and utilization through lost institutional memory and inconsistent implementation. Addressing this requires institutional commitment to stabilizing appointments through formal recognition of the role's importance and establishing systematic knowledge transfer protocols during transitions.

5.5 Developing responsive professional services

Institutions clearly expect enhanced services, including a national annual report, data integration for in-depth analysis, specialized consultation, benchmarking, best practices sharing, characteristic surveys, and deepening cooperation with the program team. Most of the services offered by Mingde program team are, in fact, being put into practice (suggesting communication gaps), and the work that has not yet fully begun is the gathering of systematic best practices collection and expanding characteristic surveys.

International benchmarks offer valuable models. NSSE exemplifies effective experience-sharing (National Survey of Student Engagement, 2026), though most extensive evaluation programs encounter a service issue with the survey of teaching and learning features (Cheng, 2011). However, Qinmin[4] and Mingde programs in mainland China can now both offer specialized services. To address the issues faced by colleges and universities from the perspective of customized services, the program team has already implemented in several institutions, including A4 University and A14 University.

6. Discussion and conclusion

This study represents the first systematic mixed-methods investigation of Mingde program's operational dynamics, marking a significant advancement in meta-evaluative research on Chinese learning assessment programs. Its pioneering status stems from three distinctive characteristics: the program's exceptional longevity and institutional participation scale, the research's commissioned nature ensuring unique access, and its comprehensive methodological approach.

The findings provide valuable dual perspectives - revealing both the program's operational strengths and implementation challenges while capturing authentic stakeholder voices. Practically, these insights can guide institutions in self-evaluation regarding results utilization and dissemination effectiveness, assess the program's actual contribution to undergraduate education quality, and inform the program team's service optimization. The study's implications extend to supporting China's ongoing undergraduate education assessment reforms, offering relevant reference value for policy and practice enhancement.

The quantitative component of this study features a limited sample size that may challenge conventional quantitative research expectations. However, sampling adequacy should be contextualized within research constraints and objectives rather than judged by absolute numerical thresholds. When investigating populations with inherently limited membership - such as the approximately 70 institutions annually participating in Mingde Program - smaller samples can still yield meaningful insights when properly contextualized. Although this research employed full-sample surveying with strong response rates (80.3%), the final 43 valid questionnaires from undergraduate institutions preclude complex statistical modeling. This limitation, however, is strategically mitigated through methodological triangulation.

The integrated qualitative dimension, comprising interviews with 25 Mingde program managers and 16 personnel from parallel programs, substantially enhances the study's explanatory power and validity. This mixed-methods design compensates for quantitative constraints by providing rich contextual understanding and diverse stakeholder perspectives. The findings maintain significant exploratory value despite sample limitations, particularly given the rare research access to established, large-scale assessment programs with over a decade of implementation history. The absence of comparable published studies, combined with the considerable challenges of securing systematic program data, further underscores the contribution of this pioneering investigation to understanding Chinese learning assessment ecosystems.

This study has several limitations. As a single-case investigation of the Mingde Program, its findings lack the comparative perspective that multi-case research could provide. Furthermore, while qualitative data were collected from university-level program managers, the study did not specifically examine managers' roles and decision-making processes during program implementation. It remains unclear how these findings might compare with other national CSL assessment programs addressing similar research questions.

Future research should therefore expand to include multiple major CSL assessment programs, such as Qinmin and Zhishan, to identify common operational patterns and contextual variations. Such comparative work would offer more robust guidance for program improvement. Additionally, focused investigation into departmental-level managers' roles across different programs could yield valuable theoretical frameworks

to support the professional development of this crucial stakeholder group.

Notes

1 Source: The 2017 National survey report of Mingde Program and the introduction of the leader of Mingde Program team.

2 Source: 2017 National Report and Regional Report from Mingde Program archives.

3 Source: The open questions from the questionnaire.

4 Source: Observation records of the Annual Meeting of the Qinmin Program on January 9, 2019.

Additional information: Author's email

2389629115@qq.com (C. Song); yrwang0213@hust.edu.cn (Y. Wang); m202575554@hust.edu.cn (C. Feng).

Funding

This work was supported by the [the Project of Huazhong University of Science and Technology Double First-class Construction] under Grant [number 2025ZKJD16].

Conflict of interest

The authors declare no conflict of interest.

References

Astin, A. W. (1985). Achieving educational excellence. Jossey-Bass.

Astin, A. W., & Antonio, A. L. (2012). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education (2nd ed.). Rowman & Littlefield.

Bao, W. (2014). Unfinished transformation: Higher education impact and student development. Educational Science Publishing House. (In Chinese)

Blaich, C., & Wise, K. (2011). From gathering to using assessment results: Lessons from the Wabash national study. National Institute for Learning Outcomes Assessment.

Bogue, E. G., & Hall, K. B. (2008). Quality and accountability in higher education (Y. Mao & L. Liu, Trans.). Beijing Normal University Press. (In Chinese)

Center for the Study of the Student Experience. (2019). CSEQ and CSXQ Survey operations closed in 2014; Consider NSSE and BCSSE.

Chang, T. (2022). Overview of student experience in the research university. In T. Chang, F. Gong, G. Lu, D. Xu, & L. Lv (Eds.), SERU in China: Advancing undergraduate learning and development. Nanjing University Press. (In Chinese)

Chen, X. (2000). Qualitative research methods and social science research. Education Science Publishing House. (In Chinese)

Cheng, X. (2011). Management of world-class universities: University management decision and higher education research. Peking University Press. (In Chinese)

Coates, H., & McCormick, A. C. (2014). Introduction: Student engagement—A window into undergraduate education. In H. Coates & A. C. McCormick (Eds.), Engaging university students: International insights from system-wide studies (pp.1-12). Springer.

Creswell, J. W., & Clark, V. L. (2017). Designing and conducting mixed methods research (2nd ed.). (Y. You & F. Chen, Trans.). Chongqing University Press. (In Chinese)

Drennan, J., O'Reilly, S., Clarke, M., & O'Brien, N. (2014). The Irish Survey of student engagement. In H. Coates & A. C. McCormick (Eds.), Engaging university students: International insights from system-wide studies (pp.109-125). Springer.

Ewell, P. T. (1985). Some implications for practice. In P. T. Ewell (Ed.), Assessing Educational Outcomes (pp. 99-110). Jossey-Bass.

Griffin, P., Coates, H., McInnis, C., & James, R. (2003). The development of an extended course experience questionnaire. Quality in Higher Education, 9(3), 259-266.

Han, J. (2012). A study on the evaluation project of National College Student Engagement Survey (NSSE) in the United States [unpublished Master's dissertation]. Capital Normal University. (In Chinese)

Joint Committee on Standards for Educational Evaluation. (2019). Program Evaluation Standards.

Klein, S. P., Kuh, G. D., Chun, M., Hamilton, L., & Shavelson, R. (2005). An approach to measuring cognitive outcomes across higher education institutions. Research in Higher Education, 46(3), 251-276.

Kuh, G. D. (2020). NSSE and NILOA: Compatriots in making outcomes assessment matter more. Assessment Update, 32(2), 10-12.

Lv, L., & Chen, S. (2010). The supporting organization of learning and teaching in the university and its experience: Research on the center for learning and teaching in Newcastle University in Australia. Comparative Education Review, 32(8), 45-50. (In Chinese)

Ministry of Education of the People's Republic of China. (2021). Notice of the Ministry of Education on printing and distributing the Implementation Plan for the Audit and Evaluation of Undergraduate Education and Teaching in Institutions (2021-2025). (In Chinese)

Muijs, D. (2004). Doing quantitative in education with SPSS. Sage Publications.

National Survey of Student Engagement. (2026). How Institutions Use Their Data and Results.

National Survey of Student Engagement. (2021). What does NSSE do?

Naughton, B. A., Suen, A. Y., & Shavelson, R. J. (2003, April). Accountability for what? Understanding the learning objectives in state higher education accountability programs. Annual meeting of the American Educational Research Association, Chicago, IL, United States.

Norrie, K., & Conway, A. (2014). Canadian perspective on student engagement. In H. Coates & A. C. McCormick (Eds.), Engaging university students: International in-

sights from system-wide studies. Springer.

Qu, Q., Fu, C., & Gao, S. (2013). Student learning survey results analysisQu of major and course situation: Take Sun Yat-sen University as an example. China University Teaching, 8, 84-86. (In Chinese)

Sanders, J. (2012). The program evaluation standards: How to assess evaluation of education program (2nd ed.). (L. Liu, Trans.). Peking University Press. (In Chinese)

Shi, J. (2016). Governing the quality: Current and future development of the student survey in China. China Higher Education Research, 2, 37-41. (In Chinese)

Shi, J., & Wang, W. (2018). Learner-Centered education for sustainable development: The academic and policy implication of research on China college student learning. Journal of East China Normal University (Educational Sciences), 36(4), 18-27. (In Chinese)

SPSSAU. (2020). Reliability analysis. (In Chinese)

Strauss, A., & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park: Sage.

Strydom, J. F., & Mentz, M. M. (2014). Student engagement in South Africa: A key to success, quality and development. In H. Coates & A. C. McCormick (Eds.), Engaging university students: International insights from system-wide studies. Springer.

The Student Survey. (2026). Frequently Asked Questions.

Wang, X., & Niu, T. (2022). Th development of American NSSE in past 20 years and its impact on Chinese college students learning. Journal of Higher Education, 43(12): 87-100. (In Chinese)

Wang, X., & Wang, J. (2018). The academic achievement assessment of Chinese university students: A two-decade review (1998-2017). Journal of Soochow University (Educational Science Edition), 6(3), 62-73. (In Chinese)

Wang, X., & Yin, H. (2022). On undergraduate education quality impact dynamics mechanism of learning assessment program—Grounded Theory analysis of China Mingde Program. Research in Higher Education of Engineering, 2, 129-135. (In Chinese)

Wei, S., & Chen, M. (2016). The influence mechanism of learning effect of undergraduates. Research in Higher Education of Engineering, 2, 167-173. (In Chinese)

Yan, F. (2010). Theoretical and practical research on educational meta-evaluation. East China Normal University Press. (In Chinese)

Yin, H., & Zheng, K. (2017). Students' course experience and engagement: An attempt to bridge two lines of research on the quality of undergraduate education. Assessment & Evaluation in Higher Education, 42(7), 1145-1158.

Zhang, J. (2014). Teaching quality evaluation and guarantee based on the value-added development of students. Beijing Normal University Press. (In Chinese)

Zhu, H., & Zhang, W. (2020). Elite college students' family backgrounds and their development: A survey on freshmen of Peking University (2016-2018). Journal of Higher Education, 41(10), 71-82. (In Chinese)