



The CIPP Model: Applications in Language Program Evaluation

Satja Sopha

Mahidol University International College

Alexander Nanni

Mahidol University International College

This article introduces the CIPP Model (Context, Input, Process, and Product) and explains its application in the evaluation of language programs. The model has long been used in various fields to evaluate programs both before they begin (e.g., by assessing the alignment of the contexts and input) and after they are complete (e.g., by evaluating how well the process has been implemented and whether the product is up to standard). The flexibility of the model is a major strength. In the field of language teaching, this model is highly relevant to curriculum development and can be applied at both the course level and the program level. This paper first introduces competing models of language program evaluation. It then introduces the CIPP Model and explains its applicability in the field of language education, providing suggestions for the application of the model. The CIPP Model has the potential to assist TESOL professionals — teachers and administrators — in improving their professional practice, curriculum design, and program evaluation. Educators in a variety of contexts would benefit from furthering their knowledge of this widely applied evaluation model.

Keywords: evaluation, CIPP Model, assessment

Introduction

The professional is, first and foremost, a bringer-about of real-world change.

(Ur, 2002, p. 388)

Professional excellence in TESOL requires lifelong learning. To best meet the needs of students and other stakeholders, language educators must continually strive to improve their professional practice. This requires innovation in response to new opportunities, adaptation of best practices from other contexts, and application of current theory in language acquisition. Additionally, educators' professional development is aided by knowledge outside of their specialization. This paper presents an example of such knowledge that is highly relevant to language education professionals: the CIPP (Context-Input-Process-Product) evaluation model. The paper first briefly introduces several models of program evaluation and explains the choice of the CIPP Model. It then explains its applicability for language program evaluation. It then provides sample questions for applying the model and suggests data collection methods.

Models of Program Evaluation

The evaluation of education programs has a long history. While Madaus and Stufflebeam (2000) date the first attempts at program evaluation to the late 18th Century, the practice expanded greatly in the 1940s, 1950s, and 1960s, when several systems of program evaluation were created. A model of evaluation should follow the principles of sound evaluation, and should therefore be “theoretically sound, user-friendly, standards-based, [and] practical” (Stufflebeam & Zhang, 2017, p. 1). While there are many rigorous models of assessment, some approaches fall into various categories of pseudo-assessment (Stufflebeam & Coryn, 2014). These include studies that only report positive results for public relations purposes, studies that are politically motivated and therefore share their results with some of the shareholders, studies that selectively draw data from only a small subset of customer reviews, and studies that are otherwise biased or incomplete in their methods. Using an established model aids assessors in engaging in rigorous evaluation and avoiding the pitfalls of pseudo-evaluation.

Each model of program evaluation has distinctive characteristics, and each emphasizes certain aspects of the program under evaluation. Stufflebeam and Zhang (2017) state clearly: “The CIPP Model is one of a number of legitimate approaches to evaluation” (p. 2). Other models of program evaluation include experimental design and the case study. Experimental design is a widely-used approach to program evaluation, including language program evaluation; however, certain limitations hinder its application in some educational contexts. The CIPP Model was created explicitly to address these limitations. Stufflebeam and Zhang (2017) list several, including the following: the lack of flexibility exhibited by experimental design and the resulting inability to react to an evolving situation; experimental design’s focus on collecting pre-treatment and post-treatment data, which removes the possibility of utilizing feedback to improve a program *during* the treatment; the requirement to have control and experimental groups, which is often impractical or impossible in educational contexts; the possibility that experimental studies find no significant difference between the control and experimental groups, in which case the study did not directly improve the program; and the focus on dependent variables at the expense of the investigation of unanticipated impacts of a program, which are often important.

Standardized tests are also commonly used to evaluate the success of educational programs, including language programs. Stufflebeam and Zhang (2017) explain the shortcomings of using standardized test results for program evaluation, including the following: the tests are not valid for assessing specific programs because they are not aligned with the program’s objectives and the students’ needs; creating a valid test to evaluate a specific project is not feasible because of the time and resources that rigorous test design requires; and standardized tests only assess students’ outcomes, not the other aspects of the project. Using standardized tests in program evaluation poses an additional danger: a project could be labeled as a failure because students failed to make gains on a standardized test when in reality the project was meeting the needs of the stakeholders.

Despite these issues, Stufflebeam and Zhang (2017) are careful to assert that while experimental design and standardized tests have limitations as sole means of evaluating programs, they “may be incorporated successfully into broader evaluation approaches” (p. 9), particularly when they are “not employed as sufficient evaluation approaches in and of themselves but as specific techniques for use in a relevant, comprehensive, fully functional evaluation approach” (p. 10). Educators using the CIPP Model can incorporate methodologies with which they are familiar; the model provides a rigorous framework that allows them to meaningfully incorporate the findings of their research to improve their respective programs.

This paper will focus on the CIPP Model for several reasons. It is extremely flexible and can encompass many other models, methods, and approaches. As Stufflebeam and Zhang (2017) explain, it “is not exclusive of other defensible approaches to evaluation or the wide range of specific inquiry techniques. Instead, it embraces and makes room for selectively and appropriately incorporating a complete pharmacopeia of sound, qualitative and quantitative inquiry methods and tools” (p. 9). The CIPP Model aligns with current language education theory in that it emphasizes both formative and

summative assessment. It is also student- and stakeholder- centric, giving voice to the many people who are involved in the program in various capacities. For these reasons, the CIPP Model is highly applicable in language teaching and learning contexts.

The CIPP Model

The CIPP Model is a “comprehensive framework for guiding formative and summative evaluations of programs, personnel, products, institutions, and systems” (Mathison, 2005, p. 60). Stufflebeam (1971) created the CIPP Model in the 1960s “to provide timely information in a systematic way for decision making, which is a proactive application of evaluation” (p. 1), later establishing the applicability of the CIPP Model for summative assessment and accountability. Unlike many other evaluation models, the CIPP Model was created for formative assessment and only later extended to summative assessment. The model is flexible by design.

The four aspects of a program evaluated by the CIPP Model are context, input, process, and product. *Context evaluations* examine “needs, problems, assets, and opportunities” (Stufflebeam, 2003, p. 2) in order to define goals and priorities. *Input evaluations* consider “alternative approaches, competing action plans, staffing plans, and budgets for their feasibility” (Stufflebeam, 2003, p. 3) in order to select a plan that will maximize effectiveness. *Process evaluations* focus on the implementation of plans. These evaluations later help to “judge program performance and interpret outcomes” (Stufflebeam, 2003, p. 3). Finally, *product evaluations* “identify and assess outcomes — intended and unintended, short term and long term” (Stufflebeam, 2003, p. 3). Identifying unintended outcomes is a key outcome of the CIPP Model. Importantly, the precise elements included in each aspect of the CIPP Model are not strictly defined. For example, although input evaluation frequently includes staffing plans, the inclusion of these plans is not an absolute requirement. The timing of the model is also flexible. It can be applied either during or after a given process.

Values are central to the CIPP Model as it is “fundamentally a values-oriented model” (Stufflebeam, 2003, p. 9). Significantly, these values should not be imposed by a single entity, e.g., by a teacher, administrator, or institution. Instead, the values driving evaluation should be co-constructed in dialogue with as broad a range of stakeholders as possible. The co-construction of the values framework is useful in that “grounding evaluations in clear, defensible values is essential to prevent evaluations from aiding and abetting morally wrong, unethical actions and instead to help assure that the evaluations will be instrumental in effectively pursuing justifiable ends” (p. 9). The use of a co-constructed values framework is a major strength of this model, as it aids in identifying previously hidden strengths and weaknesses. Some of these strengths, such as parental support or community engagement, might not be obvious if these stakeholders are not consulted. Similarly, certain weaknesses, such as indirect racism or burnout of teachers, can be brought into focus through this process (Stufflebeam, 2003).

Four standards guide the application of this model: utility, feasibility, propriety, and accuracy. All of these standards are important; however, utility is the most critical of the four (Stufflebeam, 2003). According to this standard, evaluation should be conducted when it will lead to improvement, not conducted for its own sake. In other words, evaluation should not be undertaken unless the results can be meaningfully applied. Stufflebeam (2003) states, “If a contemplated evaluation has no prospect for use, it should not be initiated, whatever its potential to satisfy accuracy, propriety, and feasibility” (p. 11). To meet the standard of feasibility, evaluation should be “realistic, prudent, viable, and frugal” (Stufflebeam, 2003, p. 10). That is, the evaluation should not be prohibitively expensive, nor should it excessively tax the resources of the evaluator or evaluating institution. To meet the standard of propriety, evaluation should be conducted “legally, ethically, and with due regard for the welfare of participants and those affected by results” (Stufflebeam, 2003, p. 10). Beyond adhering to legal and ethical requirements, evaluators must bear in mind the consequences of their actions and how the results of their assessment will impact stakeholders. To meet the standard of accuracy, evaluation must convey “technically sound

information about the features that determine the evaluand's merit, worth, probity, and/or significance" (Stufflebeam, 2003, p. 10). In this definition, *merit* refers to intrinsic value whereas *worth* refers to both intrinsic and extrinsic value. *Probity* refers to "uncompromising adherence to moral standards," and *significance* "[gauges] the reach, importance, and visibility of the enterprise's contributions and influence" (Stufflebeam, 2003, p. 10). These standards provide clear guidance as to when and how evaluation should take place.

Triangulation is essential to the CIPP Model. Evaluators will seldom have perfect access to information. They must perform their assessment based on the best information available to them. Doing so requires that they rely on triangulation, or the use of data from multiple sources to strengthen the validity of any conclusions drawn (Yin, 2014). Triangulation generally involves the collection of multiple *types* of data using a variety of methods, both qualitative and quantitative (Creswell, 2012). These sources reinforce each other, providing context and resulting in a deeper understanding of a phenomenon. The integration of data from a broad array of sources is particularly important in ensuring that stakeholders' voices are heard: "The CIPP Model requires engagement of multiple perspectives, use of a wide range of qualitative and quantitative methods, and triangulation procedures to assess and interpret a multiplicity of information" (p. 14). This is a major strength of the model.

The open-ended and flexible nature of the CIPP Model also presents a corresponding disadvantage: applying this model can be time-intensive (Anh, 2018). Evaluators must exhibit judgement in determining which types of assessment, data, and perspectives to include in their evaluation process. If they do not discriminate wisely, the evaluation process can be delayed by the inclusion of relatively inconsequential detail.

Because of its breadth, flexibility, focus on values, emphasis on utility, and incorporation of data from diverse sources, the CIPP Model is a strong addition to the professional practice of any educator, providing a formal but flexible system of evaluation which encourages educators to employ multiple forms of assessment. Stufflebeam (2003), the creator of the CIPP, operationally defines evaluation as "a process of delineating, obtaining, reporting, and applying descriptive and judgmental information ... in order to guide decision making, support accountability, disseminate effective practices, and increase understanding of the involved phenomena" (p. 10). While the CIPP Model can be applied to "programs, personnel, products, institutions and systems" (Belanger, 2006, p. 382) and is now used in "philanthropy, social programs, health professions, business, construction, and the military" (Stufflebeam, 2005, p. 31), it was originally applied in education and was designed based on the creator's experience in inner-city Chicago schools (Stufflebeam, 2016). While this section has laid out the theoretical foundations and general applications of the CIPP Model, the following section will establish the applicability of the CIPP Model in language education.

Relevance of the CIPP Model in TESOL

The CIPP Model is well suited to the field of TESOL because of its focus on formative assessment as well as summative assessment, its compatibility with the student-centered approach to education, and its potential for advocacy. Indeed, most language teachers already incorporate aspects of the CIPP Model into their practice; however, other aspects remain underdeveloped or absent. For example, Stufflebeam (2003) notes that input evaluation is "the most neglected, yet critically important type of evaluation" (p. 3). Learning about the model and formalizing evaluation methods would have the advantage of helping to fill in any gaps in evaluation procedures.

The CIPP Model's emphasis on the importance of both formative and summative evaluation is consistent with best practices in English language teaching. Evaluation is an integral part of language education. All language teaching involves some form of assessment (Brown, 2001): formal, "those designated moments during which we administer a prepared instrument to students for the purpose of measuring their language competence" (p. 402); informal, "incidental, unplanned evaluative coaching and

feedback on tasks designed to elicit performance” (p. 402); formative, “observation of the process of learning, as opposed to the product” (p. 402); or summative, “[tending] to focus on products of learning: objectively observable performance that may be evaluated somewhat independently of the process” (p. 403). No single form of assessment suffices. Formal assessments, or “what we ordinarily call tests” (Brown, 2001, p. 403), are typically present in language classes. In addition to formal assessments, though, Brown (2001) emphasizes the importance of informal, formative assessment, stating, “Our success as teachers is greatly dependent on constant informal assessment, for it gives learners information about how they are progressing toward goals and what the next step in the learning process will be” (p. 402). Regularly and systematically conducting formative assessment allows language teachers to address learning issues as they arise. Similarly, Stufflebeam (2003) emphasizes the importance of continuous assessment as “a process of continually identifying and employing appropriate means to address emergent as well as predictable and relatively fixed information needs” (p. 17). All in all, the CIPP Model is consistent with best practices in language assessment.

Student-centered education aligns with the principles of the CIPP Model in that students (and other stakeholders) are truly at the center of both processes. In student-centered learning, the satisfaction of the students’ needs is a primary concern (Clasen & Bowman, 1974). While student-centered activities are commonly used in the language classroom, truly student-centered education requires more than just student voice and choice in a limited context: “Student-centered (used in conjunction with processing, learning, or teaching) describes a learning process where much of the power during the experience resides with students” (Estes, 2004, p. 144). Students’ control over their learning is often limited, but the co-creation of values called for by the CIPP Model is a solution to this problem. Stufflebeam (2003) considers the inclusion of diverse stakeholders to be an ethical consideration. He states that “involving all levels of stakeholders is considered ethically responsible because it equitably empowers the disadvantaged as well as the advantaged to help define the appropriate evaluation questions and criteria” (p. 12). Co-creation of values gives students (as well as their parents and other stakeholders) a voice in their education, as the co-created values then inform the goals and methods in the language course program. Application of the CIPP Model of evaluation requires educators to move towards student-centered education.

Knowledge of the CIPP Model can make educators more effective change agents. Language educators who apply the CIPP Model may find themselves both more included to advocate for their other stakeholders and better equipped to do so. Many universities and schools rely on the CIPP Model for high-level decision-making (Zhang et al., 2011). By becoming conversant in the methods and terminology of the CIPP Model, educators can communicate more effectively with executives within their respective institutions. Teachers can become better advocates for the needs of their students, their students’ parents, and other key stakeholders. Program administrators can become better advocates for their teachers. The language of the CIPP Model — stakeholders, triangulation, context evaluation, etc. — can be a powerful tool in promoting stakeholders needs, acquiring necessary resources, and requesting permission to implement beneficial changes. For all of these reasons, the model is a valuable tool for any educator.

Applying the CIPP Model in TESOL

The previous sections explained the theoretical underpinnings of the CIPP Model and argued for its applicability in language education. This section will provide examples of questions that could be asked both formatively — i.e., as part of “prospective application of CIPP information to assist decision making and quality assurance” (Stufflebeam, 2003, p. 6) — and summatively, i.e., as part of “retrospective use of CIPP information to sum up the program’s merit, worth, probity, and significance” (Stufflebeam, 2003, p. 6).

The CIPP Model can be applied either by external auditors or by individuals within a program. It can be utilized at the level of an individual course — both formatively and summatively — and also at the program level. The questions in Table 1 could guide the evaluation of a language course. Stufflebeam (2003) emphasizes the role of the core values in determining goals, plans, actions, and outcomes and the corresponding evaluation, context, input, process, and product respectively. He notes, “evaluators should take into account a set of pertinent societal, institutional, program, and professional/technical values when assessing programs or other entities” (Stufflebeam, 2003, p. 8). These values inform all aspects of evaluation, and the answers to these questions could only be interpreted in relation to the core values.

TABLE 1
Questions Guiding Application of the CIPP Model at Course Level

	Formative	Summative
Context	What are the goals of the course? Which goals are our priorities?	How well did the goals and priorities align with our needs, assets, opportunities, and problems? How should we revise our goals and priorities for future iterations of the course?
Inputs	Which teaching strategies will help us achieve our goals? How does the teaching plan align with our goals?	How did this course compare to other, similar courses? How well has this course met the stakeholders’ needs?
Processes	How well do our strategies seem to be moving us towards our goals? How does our evaluative feedback inform our teaching plan?	How effective were the teaching strategies in meeting students’ needs? How did the actual teaching of the course compare to the teaching plan?
Products	How should the course be modified to better meet our goals? How can we compensate for unforeseen outcomes?	How well did our outcomes meet our goals? How can we modify our teaching strategies and teaching plan to be more effective in the future?

As mentioned, using the CIPP Model requires triangulation, or integration of data from multiple sources, to understand a phenomenon better. The sources of data in Table 2 are valuable both on their own and in combination. An educator wishing to apply the CIPP Model could complete the items most appropriate to his or her context, but using all of the data sources is not necessary. Additionally, “homemade” (Stufflebeam, 2003, p. 14) instruments, such as surveys and interview protocols, can play an important role. The more items completed, the better an educator can triangulate the performance of the course or program.

TABLE 2
Data Collection Tools for Triangulation

	Context	Input	Process	Product
Surveys of stakeholders	✓		✓	✓
Literature review	✓	✓		
Document review	✓	✓	✓	✓
Benchmarking with other courses		✓		✓
Class observations			✓	
Case studies of students			✓	✓
Experimental studies		✓		✓
Focus groups with stakeholders	✓		✓	✓
Interviews with stakeholders	✓		✓	✓

Stufflebeam (2003) provided a much more extensive list of data collection tools. This list has been adapted for language education.

Educators should collect data to the extent that it is useful and actionable. Gathering information is a vital component of evaluation, but evaluation also requires the *interpretation* of information (Brown & Abeywickrama, 2010). In other words, testing is *not* evaluation; making decisions based on test results is

evaluation (Bachman, 1990). While collecting copious amounts of data is a step towards rigorous and useful evaluation, assessment in language education should meet Stufflebeam's (2003) standard of utility, meaning it should be performed when (and only when) it can be acted on to improve a program. Evaluators should bear in mind Stufflebeam's (2003) statement that "evaluation's most important purpose is not to prove, but to improve" (p. 4).

When applying the CIPP Model of evaluation, assessors should bear in mind an important caveat. Stufflebeam and Zhang (2017) acknowledge that one of the CIPP Model's purported strengths, the engagement of stakeholders, can also be a weakness: "Excessive stakeholder engagement introduces opportunities for threatened persons to impede the evaluation" (p. 5). As the nature of CIPP is to involve stakeholders throughout the evaluation process, stakeholders who believe that the evaluation outcomes may not serve their own goals have ample opportunity to delay, obstruct, or derail the evaluation process. Arguably, however, resistance from stakeholders may also serve a purpose as may bring to light issues deserving of attention.

Conclusion

The CIPP Model is relevant to language teachers and language program administrators in a variety of contexts. The model is flexible, comprehensive, and grounded in values that are co-created among a program's stakeholders. Furthermore, it aligns well with key elements of current theory in TESOL, including an emphasis on formative assessment, student-centered learning, and advocacy. As we pursue professional excellence in the field of language education, we would do well to remember Penny Ur's (2002) statement that "the professional is, first and foremost, a bringer-about of real-world change" (p. 388). Broadening professional knowledge to encompass tools such as the CIPP Model will empower language educators to have a greater impact on the lives of their students, colleagues, and communities.

The Authors

Satja Sopha is the Senior Secretary at the Preparation Center for Languages and Mathematics at Mahidol University International College. She holds an M.A. in Public and Private Management and has over a decade of experience in language program administration.

Alexander Nanni is a senior lecturer at Mahidol University International College (MUIC), which is located in Salaya, Thailand. He began teaching English in Thailand in 2005. Alexander holds an M.Ed. in Teaching English as a Second Language from Rhode Island College and an Ed.D. in Curriculum, Teaching, Learning, and Leadership from Northeastern University in Boston.

Preparation Center for Languages and Mathematics
Mahidol University International College
999 Phutthamonthon Sai 4 Rd, Salaya, Phutthamonthon,
Nakhon Pathom 73170
Thailand

Phone: +66 880962340
Email: Alexander.nan@mahidol.ac.th
Fax: +66 0-800-2335

References

- Anh, V. T. K. (2018). Evaluation models in educational program: Strengths and weaknesses. *VNU Journal of Foreign Studies*, 34(2) 140-150.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. New York: Pearson Longman.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. New York: Pearson Longman.
- Clasen, R. E., & Bowman, R. E. (1974). Toward a student-centered learning focus inventory for junior high and middle school teachers. *The Journal of Educational Research*, 68(1), 9-11.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Upper Saddle River, N.J: Prentice Hall.
- Estes, C. A. (2004). Promoting student-centered learning in experiential education. *Journal of Experiential Education*, 27(2), 141-160.
- Madaus, G. F., & Stufflebeam, D. L. (2000). Program evaluation: A historical overview. In D. L. Stufflebeam & G. F. Madaus (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed., pp. 3-18). Kluwer Academic Publisher.
- Mathison, S. (2005). *Encyclopedia of evaluation*. London: Sage.
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. Retrieved from <http://eric.ed.gov/?id=ED062385>
- Stufflebeam, D. L. (2000). The CIPP Model for evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models* (pp. 279-317). Springer. Retrieved from http://link.springer.com/content/pdf/10.1007/0-306-47559-6_16.pdf
- Stufflebeam, D. L. (2003). *The CIPP Model for evaluation*. Presented at the Annual Conference of the Oregon Program Evaluators Network (OPEN), Portland, Oregon. Retrieved from http://link.springer.com/chapter/10.1007/978-94-010-0309-4_4
- Stufflebeam, D. L. (2007). *CIPP evaluation model checklist*. Retrieved from <http://www.wmich.edu/evalctr/checklists>.
- Stufflebeam, D. L., & Zhang, G. (2017). *The CIPP evaluation model: How to evaluate for improvement and accountability*. Guilford Publications.
- Stufflebeam, D. L., & Coryn, C. L. (2014). *Evaluation theory, models, and applications*. John Wiley & Sons.
- Ur, P. (2002). The English teacher as professional. In C. R. Jack & A. R. Willy (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 388-392). Cambridge University Press.
- Yin, R. K. (2014). *Case study research: Design and methods*. London: Sage Publications.
- Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011). Using the context, input, process, and product evaluation model (CIPP) as a comprehensive framework to guide the planning, implementation, and assessment of service-learning programs. *Journal of Higher Education Outreach and Engagement*, 15(4), 57-84.