PERGAMON

Evaluation and Program Planning 22 (1999) 1–11

EVALUATION and PROGRAM PLANNING

A Bayesian model to predict the success of the implementation of health and education innovations in school-centered programs

K. Bosworth[a,*], P.M. Gingiss[b], S. Potthoff[c], C. Roberts-Gray[d]

[a] College of Education, University of Arizona, Tucson, AZ 85721, USA
[b] University of Houston
[c] University of Minnesota
[d] The Resource Network, Austin, TX

Received 22 July 1995; received in revised form 18 January 1997; accepted 24 May 1998

Abstract

Health and education practitioners, evaluators, and researchers have little guidance to help them translate implementation research into meaningful implementation strategies. This article describes the development and testing of a model to help schools assess their likelihood of successfully implementing health education innovations. The model was developed using an integrative group process technique that captures experts' qualitative and quantitative judgments as a subjective Bayesian probability model. The experts developed a measurable definition of successful implementation, identified eight factors containing 40 questions relevant for predicting successful implementation, and specified the diagnostic value of each of the factors. Internal validation showed a correlation of 0.92 between the model scores and the experts' direct ratings of 100 hypothetical school profiles. Preliminary external and content validation have been conducted. Application of the model to planning, management, and evaluation of school-based innovations is discussed.

1. Introduction

As the complexity of factors influencing the health and well-being of children today is increasingly recognized, the programmatic solutions also are becoming more complex. Evaluation reports of program implementation note numerous implementation barriers and failures at the organizational, work group, and individual levels.

Implementation is considered to occur when an innovation is put to use (Rogers, 1983). Thus, the implementation stage begins after the adoption decision is made and culminates when the innovation 'disappears' either because it has become so thoroughly integrated into everyday practices that it is no longer visible as an innovation or because it has been discontinued (Yin, 1979; Scheirer, 1990). An innovation is an idea, practice, service, technology or other object that is perceived as new by an individual or other unit of adoption (Rogers, 1983). The innovation may include microcomputer software, videodisc, new instructional strategies or programs, new policies and guidelines, and/or action plans for comprehensive, collaborative programming.

Over time, various perspectives for predicting implementation success have been provided. Much of the early literature focused on the adoption phase of the diffusion process (Rogers & Shoemaker, 1971). In this literature, use of an innovation was viewed as an event, rather than a phenomenon that has a time line and process (Hall, 1992). Other areas of research and theory that shaped perspectives on implementation included work on organizational development (Hernandez & Kaluzny, 1994; Rothman, Erlich & Teresa, 1981; Scheirer, 1990; Yin, 1979); major school-based studies such as the Rand Change Agent Study (McLaughlin, 1990); a study of the effects of the National Diffusion Network (Emrick, Peterson & Agarwala-Rogers, 1977) and a study of the school improvement processes in response to innovations (Huberman & Miles, 1984). Another major set of educational implementation studies focused on schools, individual teachers, and classroom practices (Hall & Hord, 1984). In the health promotion literature, a number of related factors critical to successful implementation have been identified (Bosworth & Yoast, 1991; Brink, Gingiss & Gottlieb, 1991; Gingiss & Engel, 1995; Hall & Loucks, 1977; Levenson-Gingiss & Hamilton, 1989; Ottoson & Green, 1987; Roberts-Gray & Scheirer, 1988). Additionally, in the early 1990s the National Cancer Institute funded two national centers to research the diffusion of

* Corresponding author. Tel.: 520 626 4350; fax: 520 626 6005; e-mail: Bosworth@u.arizona.edu

0149-7189/99/$ - see front matter
PII: S0149-7189(98)00035-4

health promotion programs and policies in schools. One of these programs based its intervention on diffusion theory and social learning theory, while the other used organizational development theory (Goodman & Steckler, 1990; Parcel et al., 1989). Related reports of the implementation of innovative, complex, multi-component, collaborative programs which integrate schools, health and social service programs, have indicated numerous issues which span delivery system boundaries (Crowson & Boyd, 1993; Knapp, 1995; Marks & Marzke, 1994).

A review of implementation literature confirmed that substantially over 300 variables can be identified that might be integrated into a model to guide implementation planning and monitoring (Roberts-Gray, 1985). In spite of the existing research, health and education practitioners, evaluators, and researchers are left with little guidance to help them translate this research into meaningful strategies to improve implementation efforts. Without guidelines, it is very difficult to develop focused staff development and technical assistance programs to strengthen implementation. Given the diminishing resources available to health and education sectors, efficient utilization of these limited resources for program support is critical.

A solution is to translate available implementation research into a model to assist schools and community groups providing school-based or school-linked programs to plan for and monitor key processes. The goal of such a model would be to provide guidance to strengthen implementation efforts to improve the likelihood of success. Unfortunately, an extensive, convergent data base upon which to build a predictive model of successful implementation is lacking in the field. However, a number of individual researchers and practitioners are intensely involved in the research and practice of improving the implementation stage of planned change. The approach taken in this research was to elicit the knowledge of an expert panel of implementation researchers and practitioners in the form of a mathematical model. The model delineates the factors the panelists defined as important to assess the implementation process, and the diagnostic strength of each factor in contributing to the likelihood of successful implementation. This article describes the theoretical background for this approach, and the development, testing, and application of the model in assessing and guiding implementation of health innovations in schools.

2. Methodology

2.1. Background

Modeling human judgment has been central to the field of decision analysis (von Winterfeldt & Edwards, 1986). Decision analysis provides methods for eliciting people's values or preferences and mapping them into numbers to quantify their judgments (Gustafson, Cats-Baril & Alemi, 1992). The elicitation process involves assessing two types of knowledge: (1) the attributes or factors the person considers important when making a judgment, and ways to describe levels (e.g., high, medium, low) for each of those factors; and (2) the subjective weight or importance of each factor level. A scenario or option under consideration is evaluated by assessing how it measures (e.g., high, medium, low) on each factor level, with the corresponding weights for those levels combined mathematically to arrive at an overall numerical score for that scenario.

The mathematical model can support decision making in a number of ways. In this research, the model predicts the likelihood of a school successfully implementing an innovation. The model calculates a numerical score that represents the probability of successful implementation given that school's characteristics. School personnel can use this overall score to see how likely they are to achieve success. They can also use the model as a diagnostic tool by looking at how they rate on each of the factors of the model to see where they are experiencing difficulties. The model also provides the capability to look at how improving in areas where they are having difficulties would improve their overall likelihood of success. Although such models can be used in paper and pencil format, computerization greatly enhances their usability. This decision analytic approach has been used to model expert judgment in a variety of areas, including evaluating the quality of psychiatric emergency care (Gustafson, Sainfort, Johnson & Sateia, 1993), assessing school drug prevention programs (Bosworth & Yoast, 1991), measuring the severity of trauma injury (Fryback & Keeney, 1983), and evaluating the quality of long term care (Gustafson et al., 1990).

The decision analytic models generally take one of two mathematical forms: a subjective Bayesian model, or a multiattribute utility model (von Winterfeldt & Edwards, 1986). Subjective Bayesian models are grounded in Bayesian statistics, which uses Bayes' theorem to revise one's opinion about the probability of an outcome in light of new information (von Winterfeldt & Edwards, 1986). This approach ascribes to the normative view that subjective probabilities elicited from people describe orderly opinions about how one's beliefs should change when given new evidence. Subjective Bayesian models multiply together all the data that are available. If a piece of data is missing, it is simply omitted from the Bayes' theorem calculation. Thus, subjective Bayesian models are not very sensitive to small amounts of missing data. Multiattribute utility models also capture subjective information to guide decision making. However, these models rely on weighting attributes and creating utility functions across the levels of each attribute. They suffer when data are missing, requiring re-weighting all the remaining attributes to arrive at a score (Gustafson, Sainfort, Johnson & Sateia, 1993). The subjective Bayesian model was chosen as the mathematical form for this research because its framework better fits the prediction task at hand, and it is easier to use when small amounts of missing information may occur.

2.2. *Subjective Bayesian models*

The purpose of the Bayesian model is to predict the probability of successful implementation of a health innovation by a school. The overall probability calculated by the Bayesian model ranges from 0 (almost no chance of successful implementation) to 100 (almost certain chance of successful implementation). The format of Bayesian models requires that there be competing outcomes to be predicted that are mutually exclusive (only one of the outcomes can happen) and exhaustive (the probabilities of the outcomes must sum to 1). Unless there is clear reason to the contrary, one should limit the number of outcomes being forecasted to two to simplify the model building process (Gustafson, Cats-Baril & Alemi 1992). Thus, in our model the two competing outcomes were successful implementation and failed implementation.

Bayesian models contain three components, which will be discussed specific to our model: prior odds of success, likelihood ratios for each level of each factor, and posterior odds of success. The mathematical form of the Bayesian model is shown below, with the example assuming there are two factors that are used to predict successful implementation.

$$\frac{P(\text{Success}|\text{Fact1}_b, \text{Fact2}_a)}{P(\text{Failure}|\text{Fact1}_b, \text{Fact2}_a)} = \frac{P(\text{Fact1}_b|\text{Success})}{P(\text{Fact1}_a, \text{Failure})} \times \frac{P(\text{Fact2}_a|\text{Success})}{P(\text{Fact2}_{ai}|\text{Failure})} \times \frac{P(\text{Success})}{P(\text{Failure})}$$

posterior odds of success = likelihood ratio for level 'b' of Factor 1

× likelihood ratio for level 'a' of Factor 2

× prior odds of success

The prior odds of success is the probability of success divided by the probability of failure, before one has any information about the characteristics of the school. A non-informative prior was assumed (e.g., no a priori belief about the probability of success was assumed before collecting information about the school). This assumption can be as accurate in predictive ability as assuming prior knowledge (Gustafson, Kestly, Greist & Jensen, 1971). For any given school, each factor will be present at only one of several levels. The likelihood ratio for each level of each factor indicates the 'weight' for that piece of data about the school. The posterior odds of success is calculated by multiplying the prior odds (which in our model is 1/1 for a non-informative prior) by each of the likelihood ratios. The final probability score of success can be calculated by:

Probability of successful implementation

= (Posterior odds)/(1 + Posterior odds)

This form of Bayes' theorem requires that the factors be conditionally independent (von Winterfeldt & Edwards, 1986). That is, if one is told that a school successfully implemented an innovation, knowing that a school was at a given level on one factor would not provide any information about the school's performance on any other factor. Conditional independence was assessed by the experts during the model-building process, and all factors were identified as being conditionally independent. This greatly simplifies the likelihood ratio estimation process described in Step 7 below. If two factors are not conditionally independent, the expert panel would be required to estimate more complex likelihood ratios. For example, suppose Factor 1 is not conditionally independent of Factor 2. Rather than simply estimating the probability of Factor 1 being at level b given that success occurred, the panel would have to estimate the probability of Factor 1 being at level b given that success occurred *and* Factor 2 is at level a, b, or c. This increases the complexity of the likelihood estimation process and the number of estimates that must be elicited.

2.3. *Integrative group process*

The elicitation process used to model the panelists' expertise in the form of a subjective Bayesian model is known as an integrative group process technique (Gustafson et al., 1992). The expert panel approach used is an economical and reliable approach for developing prediction models. Previous studies indicate that mathematical models developed using this approach can be as accurate in prediction as those developed using more laborious, time-consuming, and costly approaches that use extensive empirical records and primary data collection (Gustafson et al., 1971). An integrative group process approach was particularly appropriate in this research since the goal was to develop a usable predictive model in a short time frame. A convergent data base upon which to build a predictive model of successful implementation does not exist, necessitating the use of expert estimation.

Several tasks are necessary in creating a subjective Bayesian model: (1) establish a measurable definition of successful implementation, (2) identify a set of relevant factors for predicting successful implementation, (3) specify the likelihood ratios for each factor level in predicting successful implementation, and (4) test the reliability and validity of the model. The model was developed through

Table 1
Steps in the model development process

Step	Description	Time
1	Selection of panel members (experts)	2 months pre-meeting
2	In-depth telephone interviews with panelists	2 weeks pre-meeting
3	Development of the straw model	1 week pre-meeting
4	Defining implementation success	Day 1
5	Identifying the factors	Day 1
6	Developing measures for the factors	Day 1
7	Developing likelihood ratios	Day 2
8	Rate hypothetical profiles	Day 2
9	Re-rate hypothetical profiles	1 month post-meeting
10	Test internal validity of model	Post meeting

a ten-step process which allowed panelists to complete the above tasks. The ten-step process used was based on extensive research by Gustafson et al. (1992). Each step is identified in Table 1 and described in this section.

2.4. *Step 1: selection of panel members*

Previous experience suggests that the panel should be no more than five or six members and conducted on a very tight time line of approximately eight to twelve hours of meeting time distributed across two consecutive days (Gustafson Bosworth, Treece, Moberg & Hawkins, 1992; Gustafson, Tianen & Greist, 1981). Criteria for selecting panel members included persons who have theoretical knowledge of the implementation process through research, are knowledgeable about the various kinds of new programs and technologies for school-centered health and education programs, and have had first-hand experience as 'change facilitators' in schools. Panelists represented theoretical backgrounds and expertise in social psychology, school change, organization change, public health education, and diffusion research.

2.5. *Step 2: in-depth interviews with panelists*

In telephone interviews conducted about two weeks before the meeting, the individual panel members were asked to give their definitions of implementation success of school-centered health programs, discuss the two or three factors that they perceive to be most useful in predicting success, and then to explore other factors that they think should be included in a model for predicting implementation success.

2.6. *Step 3: development of the straw model*

The definition of success and the list of factors elicited from each panel member in Step 2 were used to draft a preliminary 'straw' model. This model included all factors that appeared on the list for more than one member and at least one factor from the list generated by each individual. The straw model was used at the start of the panel meeting to initiate discussion.

2.7. *Step 4: defining implementation success*

Although the term 'success' in reality may be defined on a continuum, in an ideal sense it can be thought of as a one-time future event: implementation either is or is not successful. Given this guideline and a working definition developed from data provided in their individual interviews, the panelists collaboratively developed a final definition. The precision of the definition is important, so that all panelists are working with the same criteria for program success.

2.8. *Step 5: identifying the factors*

The panel's next task was to collaboratively develop a list of measurable, conditionally independent factors to be included in the mathematical model based on their collective theoretical and practical expertise. The panel first reviewed the straw model, and through extensive discussion, added missing items and combined duplicates. In developing the final list of factors, panelists considered two criteria. First, the factors needed to be conditionally independent of each other. To determine conditional independence, the panelists were asked to assume that a school had been successful (or had failed). They were then told that the school had responded that it was at level 'a' on factor 1. They were then asked whether that piece of information told them 'a lot' about how the school might have responded to any of the other factors. This was repeated for numerous combinations until the panelists indicated that there were no sets of responses that would indicate such redundancy. Thus, the assumption of conditional independence was met for all factors. If it had not been, the option would have been to omit the conditionally dependent factors, or elicit more complicated likelihood ratios that did not assume conditional independence among the factors. The second criterion in determining the final set of factors was that they should reflect information easily accessible to school personnel, i.e., personnel at the school should have access to information that would enable them to assign a 'score' for the factor.

2.9. Step 6: developing measures for the factors

Next, the panel was asked to identify key components of each factor and to develop operational definitions for each level. Seven of the eight factors consisted of a number of components. As part of their discussion, the panel developed operational definitions for each level and scales for each component. They then specified rules to aggregate responses to items in each component into an overall three-point scale (high, medium, or low level) for that factor. When the panel had agreed on the factors, their measures, and their levels, Day 1 was concluded.

2.10. Step 7: developing likelihood ratios

Day 2 was devoted to eliciting numerical judgments from the panel needed to develop the likelihood ratios for the subjective Bayesian model. These likelihood ratios expressed the 'weight' of each factor level in contributing to the likelihood of successful implementation at the end of three years. This process involved eliciting the experts' estimation of the relative diagnostic value of each factor level. To develop likelihood ratios, panel members were asked to consider 100 hypothetical schools where a health or education program was known to have been successfully implemented. Looking at only one factor at a time, the panel 'scored' the factors for the 100 schools. For example, of 100 schools where implementation was a success, 70 might have had high compatibility and 30 might not. The panel then considered 100 hypothetical schools where implementation was not successful and performed the same estimation. In estimating the likelihood ratios, an Estimate-Talk-Estimate (ETE) procedure was used (Gustafson et al., 1992). Panel members were asked to make estimates individually of the predictive power of each factor. Each member clarified and justified his or her responses and definitions to the panel when the estimates differed. Finally, panelists re-estimated again individually. Consensus is not required. Remaining differences between panel members were resolved by equally weighting each expert's likelihood ratio estimate, using the geometric mean of the estimates as the final estimate of diagnostic value (Rowse, Gustafson & Ludke, 1974).

2.11. Step 8: rate hypothetical profiles

To provide data with which to test the model's internal validity, panel members were asked to make subjective estimates of the probability of successful implementation for hypothetical school profiles representing the eight factors in the model. Profiles were randomly generated from the eight factors in the model, i.e., random numbers were computer generated to specify scores of 1 (low), 2 (medium), or 3 (high) for each of the factors to generate factor profiles for 100 hypothetical schools. The panel then estimated the probability of success for each profile by assigning a number on a scale of 0 to 100, where *zero* means absolutely no probability of successful implementation, and 100 means 100% probability of successful implementation. These estimates, called 'holistic ratings', ask the panelists to consider all the factors in total and provide an overall estimate of the probability of successful implementation. The ETE procedure enabled the individual panel members to discuss their holistic ratings of each hypothetical profile, and re-rate the profiles based on the resulting discussions. A final group estimate for each profile was obtained by averaging the individual estimates. When they completed this task, the panel's meeting came to an end. The internal validity of the model was determined by comparing the holistic estimate of each profile to the score calculated by the model for each profile. This is discussed further in Step 10 and in the results section.

2.12. Step 9: re-rate hypothetical profiles

Approximately one month after the group meeting, the panel received a mailed sample of 40 of the original hypothetical school profiles. They were asked to again make subjective estimates of the probability of successful implementation for each of the profiles (a repeat of Step 8). By completing this activity, the panel supplied data to test the stability of the holistic ratings and evaluate reliability.

2.13. Step 10: test internal validity of model

Testing internal validity involved correlating the group holistic rating for each profile to the corresponding score generated for that profile by the Bayesian model. The higher the correlation, the better the model is capturing the group's expert judgment. Also, it was investigated whether all the factors should be included in the model, or only those with the most diagnostic power. The diagnostic power of a factor refers to the range between its largest and smallest likelihood ratio corresponding to the 'best' and 'worst' level for that factor. For example, Factor 1's largest likelihood ratio, e.g., the likelihood ratio that most pulls the probability towards the 'success' end if the school is at that level on Factor 1, is 2.57 | (Table 2). Its smallest likelihood ratio, on the other hand, is 1/2.44, which would pull a school towards the 'failure' end if it is at that level on the factor. This range of $2.57+2.44=5.01$ provides a heuristic for judging the

Table 2
Model factors, components, likelihood ratios, and diagnostic power

Factor	Components	Likelihood ratios[1]	Diagnostic power[2]
1. *Facilitation Process*	Planning Training Coaching Monitoring Communicating	High: 2.57/1 Medium: 1.29/1 Low: 1/2.44	5.01
2. *Resources*	Materials Staff Funds Daily time Implementation time frame Facilities	High: 2.18/1 Medium: 1/1.03 Low: 1/2.11	4.29
3. *School-Based Leadership*	Principal Program leaders Team structure & function Priority	High: 2.07/1 Medium: 1/1.01 Low: 1/2.16	4.23
4. *Implementers*	Professional preparation Commitment to health Implementation skills Enthusiasm Perceptions of role compatibility Perceptions of relative advantage	High: 1.90/1 Medium: 1.09/1 Low: 1/2.23	4.13
5. *External Environment*	Turbulence Support Opposition Guidelines Bureaucracy	High: 1.70/1 Medium: 1.13/1 Low: 1/1.65	3.35
6. *Compatibility*	Priorities Structure Student needs Culture Past success	High: 1.56/1 Medium: 1.11/1 Low: 1/1.60	3.16
7. *External Leadership*		Active: 1.60/1 Somewhat Active: 1.11/1 Not Active: 1/1.25 Not Identified: 1/1.52	3.12
8. *Innovation Characteristics*	Complexity Relative advantage Ease of use	High: 1.32/1 Medium: 1.15/1 Low: 1/1.52	2.84

[1] Factors with multiple components contain phrase anchored scales with rules generated by the expert panel to classify responses into high, medium, or low levels for the factor overall.
[2] Diagnostic power is the range between the largest and smallest likelihood ratio for each factor (e.g., 2.57 + 2.44 = 5.01).

amount of information that Factor 1 has the potential to give you compared to other factors. Suppose there was a factor whose largest likelihood ratio was 1.1/1, and whose smallest likelihood ratio was 1/1.1. Neither of these likelihood ratios is very informative, and the diagnostic power of that factor would be very small.

3. Results

This section describes the components of the model including the definition of success, the predictive factors and their likelihood ratios, and the validity of the model.

3.1. Measurable definition of successful implementation

Based on data provided from individual interviews and facilitated group discussion among the panel members, (Step 2), a measurable definition of successful implementation of school-based health programs was developed by the panelists:

> After three years the critical components are used with 65% fidelity to the original design with implementer buy-in by 80% of eligible implementers and received by 80% of consumers.

The panel, after much discussion, determined that three years was the amount of time required for programs to move from adoption and initial trials to more mature levels of implementation. 'Eligible implementers' include teachers, counselors, nurses, coaches, and other direct-service personnel involved in school-centered programs. 'Consumers' are the students and anyone else who is in the target audience for the particular innovation.

3.2. *Factors relevant for prediction of successful implementation.*

Eight factors were identified: Facilitation Process, Resources, School-based Leadership, Implementers, External Environment, External Leadership, Compatibility of Innovation with the Setting, and Innovation Characteristics. Table 2 includes these factors and their components, in descending order of the diagnostic power of each factor (see Step 7). As noted, seven of the eight factors consisted of a number of components. Five-point, phrase-anchored scales were generated for each component. Rules to aggregate the levels on each factor's components into an overall three-point scale (high, medium, or low level) were agreed upon. For example, to achieve a score of 'High' on the Facilitation Process factor, a school would have to achieve at least a 3.5 or better (on the 5-point scale) for each of the components of the facilitation process (e.g., Planning, Training, Coaching, Monitoring, and Communicating). The only exception to this was the factor 'External Leadership', which had no components, but rather was one four-point, phrase-anchored scale.

Factor 1, *the facilitation process*, has the most diagnostic power. Components of the factor pertain to (a) the presence of a written plan, (b) provision for on-going training for staff and faculty, (c) the level of technical assistance and coaching planned, (d) the frequency and quality of monitoring and feedback provided by colleagues or administrators, and (e) how many communication channels exist for regular two-way communication.

Factor 2, *resources*, addresses the adequacy of a variety of resources, including (a) materials and supplies, (b) staffing, (c) funding, (d) daily time allocated for planning and follow-through, (e) longer-range time for planning and trying out the innovation, and facilities.

Factor 3, *school-based leadership*, covers (a) level of engagement of the principal (work group leader) with the innovation, (b) identification of and administrative commitment to the program leader, (c) working relationship between the principal and the program leader, and (d) the relative importance of the program to the principal.

Factor 4, *implementer characteristics*, considers the following characteristics of teachers or other direct service providers: (a) level of professional preparation and experience, (b) commitment to student health, (c) implementer capacity to meet innovation implementation requirements (level of skill and experience), (d) commitment to implementation of the innovation, (e) compatibility of implementation requirements with other job expectations and responsibilities, (f) compatibility of implementation requirements with personal definitions of role and professional identity, and (g) belief in the relative advantage of the innovation over current practice.

Factor 5, *external environment*, pertains to (a) level of disruption or turmoil outside the school, (b) level of support from parents and community, (c) level of community opposition, (d) support for use of the innovation through federal, state, and district policies and guidelines, and (e) level of bureaucratic hurdles at the school level.

Factor 6, *compatibility*, examines how compatible the innovation is with the following: (a) school priorities, (b) school administrative structure, (c) student needs, (d) school culture, and (e) school history of implementing innovations.

Factor 7, *external leadership*, addresses the level of engagement of a person at the district (organizational) level to facilitate, coordinate, support, and advocate for the innovation. This factor is unique in that it contains only one item and uses a four-point scale.

Factor 8, *innovation characteristics*, pertains to the following: (a) complexity of the innovation, (b) relative advantage over what is currently being done, and (c) ease of implementation.

Figure 1 contains an example of how one would use the model to calculate a probability score of successful implementation based on the characteristics of a hypothetical school. In this example, school personnel would rate themselves on the five-point phrase anchored scales on each of the seven factors with multiple components, and aggregate their performance on the components into the high, medium, or low level for each factor using the criteria identified by the expert panel. For the factor that did not have multiple components (external leadership) they would choose the phrase that best matched their situation. They would then multiply the likelihood ratios for each factor level relevant to their profile by the prior odds to arrive at overall odds of successful implementation. This example shows that the school has a little better than even chance of being successful given its profile. If the school improved its facilitation process to the highest level, however, its probability of success would rise to 70%

3.3. *Internal validity and reliability*

The first test of the model's internal validity was to determine how well its scores correlate with the subjective estimates of panelists on hypothetical profiles of schools (Step 10). For the 100 profiles, the panel members' direct estimates (e.g., a direct holistic rating on a 0–100 scale)

FACTOR	LEVEL	LIKELIHOOD RATIO
Facilitation Process	Medium	1.29 / 1
Resources	High	2.18 / 1
School-based Leadership	Medium	1 / 1.01
Implementers	Low	1 / 2.23
External Environment	Medium	1.13 / 1
Compatibility	Low	1 / 1.60
External Leadership	Somewhat Active	1.11 / 1
Innovation Characteristics	Medium	1.15 / 1

Multiplying the prior odds (1/1) and each likelihood ratio corresponding to each factor level for this school profile yields a posterior odds of 4.06 / 3.60 or 1.13 / 1. This is a 53% probability of successful implementation (or little better than a 50-50 chance of success).

Fig. 1. Example calculation based on hypothetical school profile. Panelists selected the scale of Active (A), Somewhat active (S), Not active (Not), and Not identified (No) for this factor.

for each profile were compared to the values resulting from the application of the Bayesian model. This comparison demonstrated a correlation of 0.92 ($P < 0.000$), indicating a high level of agreement between the model and the experts' subjective estimates. The model also performed at a high level of agreement with the subjective holistic profile ratings of each of the individual panel members (i.e., 0.80, 0.081, 0.70, 0.85, and 0.82). The test retest reliability of the holistic ratings was also high, yielding a correlation of 0.80 for the initial ratings of panel members during their group meeting and additional ratings they made approximately one month after the meeting. The correlation between the model scores and the subjective ratings at a second point in time was 0.93. These results give confidence that the model accurately captured the panelists' expert judgments.

The model containing all eight factors better captured the panelists' holistic profile ratings than did a model containing only the four most predictive factors. When only the four largest weight factors (Facilitation Process, Resources, School-based Leadership, and Implementers) were included in the model, the correlation dropped to 0.83. These results indicate that all eight factors should be included in the model.

4. External model testing

In the development of the model, a high degree of consensus among panel members was reached, even though the panelists came from different perspectives and backgrounds and did not know each other prior to the meeting. Analysis indicates that the model accurately captures their collective expertise. Important next steps are to examine external validity and content validity. Since the examples of school profiles used during the model development process were hypothetical, tests of content validity and external validity across a variety of innovations on a sample of actual schools are underway.

4.1. Follow-up tests for content and external validity

The first round of school-based health centers funded by the Texas Department of Health, in cooperation with Texas Education Agency, have been sites for initial testing of the model during the start-up phase of project operation. Eleven elementary schools and 10 secondary schools during their first year of funding participated. Sixty school administrators, direct service providers and community or district-level administrators completed the questionnaire. In addition, follow-up interviews were conducted using semi-structured interviews with 61 key individuals involved in campus-level operation of the health centers to ascertain content validity of the model and determine additional factors influencing implementation. The level of content validity was extremely high. The model was found to be fast and easy to use, and the benefits for staff development, technical assistance, and process evaluation were ascertained. Particularly, the heavy weighting of the most diagnostic factor, the facilitation process, was reflected in participant perceptions of the foremost barriers to implementation. Scores from the implementation model were found to be an effective and meaningful way to compare districts and schools within districts to provide the state with a basis for prioritizing scarce training and technical assistance resources. (Ginsiss & Engel, 1995). At the end of the first three years of program operation, follow-up assessment of implementation levels will be assessed to examine predictive powers of the factors.

Plans are currently underway to construct case studies from several national school-based diffusion efforts in health and education using mature programs which have pre-established criteria for successful implementation. These case studies, with rich contextual backgrounds, will then be presented to another expert panel for rating to assess the predictive values of the factors on 'real' cases.

The model is currently being used in evaluation projects. The Texas affiliate of the American Heart Association is using it to assess the value of various program components. The model has been applied to the implementation of peer mediation programs in 12 elementary schools in Indiana as an evaluation tool. With support from the Attorney General's Office and the Indiana State Bar Association, teams from these elementary schools have been trained and are charged to implement a peer mediation program. The model has assessed their baseline implementation plan and will be included in the follow-up evaluation.

5. Discussion

Although much research exists on implementation of innovations, a gap exists in being able to translate this research into meaningful implementation strategies for school-based health program innovations. This project helped to bridge this gap by modeling expertise of implementation researchers and practitioners. Their expertise is represented in a subjective Bayesian model, which can be used by schools to guide and evaluate their implementation processes. The panelists identified and defined the components of eight factors as critical to successful implementation: facilitation process, resources, school-based leadership, implementers, external environment, external leadership, compatibility of innovation, and characteristics of the innovation.

The panel defined a standard for implementation success: that within three years, users should have achieved 65% fidelity to the original design, the innovation would be used by 80% of available implementers, and would reach 80% of the designated consumers. This definition helped ensure that all panelists were working with the same definition when they were building the model, although in practice the users of the model may choose to define their own performance criteria. The standards for implementation, however, provide guidelines for programs when establishing time lines for measuring implementation success, as well as levels of completeness, fidelity, and penetration. Defining critical components of an innovation, establishing criteria for measuring success, and providing evaluation tools to measure these are a critical part of the process for planning for implementation.

The implementation model showed high internal and external validity. It captured the panelists' expertise and demonstrated its utility as an implementation guide for actual schools in field tests. The model has proved to be an excellent self-assessment and planning tool to guide people charged with implementing innovations through these initial processes. The model may be used at various phases of program implementation.

1. During program development. An optimal initial use of the model is during the developmental process. As developers generate innovations, the model provides a tool for testing those new concepts and approaches in the field by providing feedback on intended users' strengths and barriers. With this information, developers and distributors can forecast the accessibility, usability, and fit of the product with settings of intended use in time to make crucial product modifications. Up-front needs for product training and technical assistance can be identified during this phase so that resources can be identified to accompany the innovation.
2. Before implementation. The model may be used before implementation, when a new program has been adopted or is being considered for adoption and school or agency personnel wish to predict the likelihood of making an innovation work in their setting. Use at this time allows adopting organizations to systematically plan for organizational, work group, and user changes necessary to maximize implementation success. Since the system is focused on the campus level, scores across campuses allow a school district or program coordinator to compare campus capabilities to plan for allocation of assistance.
3. After completion of implementation trials. The model may be used after initial efforts to use the program have been underway. Use at this point allows the user to diagnose how the new program is working and to identify areas which require further attention. After the new program has been used repeatedly, the model may be used to determine barriers and strengths which may influence program institutionalization.
4. Tracking program changes in delivery capacity. When used before implementation as a needs assessment, program data provide quantitative measurement of factors which may assist and impede program use. These values may then be used as criterion measures for the development of site-specific objectives.

At various phases during program implementation the same measures may be repeated to assess changes in program delivery capacity. This is an especially effective way to assess the impact of follow-up staff development and technical assistance programs. Results may also be used to explain program implementation successes and failures.

The next task for improving this model is to design a strategy to help turn this model into a tool that can be more easily used in the implementation process. Computer technology could assist in making the model easily accessible to schools. Using an interactive computer assessment interview, a school representative could assess that school on each factor. Strengths and weaknesses could be identified. Information and resources to boost scores on the factors could be presented. This concept is similar to that which guided the development of DIADS (Drug Information Assessment and Decision Support), a computer-based system designed to help schools plan effective drug abuse prevention programs (Bosworth & Yoast, 1991). This system combines an assessment function based on a subjective Bayesian model with information, resource data bases, and action plans to guide the selection, development, and implementation of school-based drug prevention programs.

In summary, this model has proved to be a useful tool to assist schools in implementing health and education innovations. Computerizing the model and providing on-line information and resources regarding implementation via a 'friendly' interactive computer program would enhance its usability.

Acknowledgement

The project was funded by the National Cancer Institute (Grant No. 1 R43 CA57096-01).

References

Bosworth, K., & Yoast, R. (1991). DIADS: Computer-based system for development of school drug prevention programs. *Journal of Drug Education, 21*(3), 231–245.

Brink, S., Gingiss, P., & Gottlieb, N. (1991). An evaluation of the effectiveness of a planned diffusion process. *Health Education Research, 6*(3), 183–191.

Crowson, R. L., & Boyd, W. L. (1993, February). Coordinated services for children: Designing arks for storms and seas unknown. *American Journal of Education, 101*(2), 140–179.

Emrick, J. A., Peterson, S. M. & Agarwala-Rogers, R. (1977). *Evaluation of the national diffusion network: Findings and recommendations* (Vols. I and II). Menlo Park, CA: Stanford Research Institute.

Fryback, D. G., & Keeney, R. L. (1983). Constructing a complex judgmental model: an index of trauma severity. *Management Science, 29*, 869–883.

Gingiss, P., & Engel, M. (1995, October). *Factors influencing the likelihood of successful implementation of school health centers.* Paper presented at the annual meeting of the American School Health Association, Milwaukee, WI.

Goodman, R. & Steckler, A. (1990). Mobilizing organizations for health enhancement: Theories of organizational change. In K. Glanz, F. Lewis, & B. Rimer, (Eds.). *Health behavior and health education: Theory, research and practice* (pp. 314–341). San Francisco, CA: Jossey-Bass.

Gustafson, D. H., Bosworth, K., Treece, C., Moberg, D. P., & Hawkins, R. (1992). Drug abuse prediction and prevention in high risk families. SARA Project: Final Report to NIDA (National Institute for Drug Abuse. Center for Health Systems Research & Analysis, University of Wisconsin-Madison.

Gustafson, D. H., Cats-Baril, W., & Alemi, F. (1992). *Systems to support health policy analysis: Theories, models and uses.* Ann Arbor, MI: Health Administration Press.

Gustafson, D. H., Kestly, J., Greist, J., & Jensen, N. (1971). Initial evaluation of a subjective Bayesian diagnostic system. *Health Services Research, 6*, 204–213.

Gustafson, D. H., Sainfort, F. C., Johnson S. W., & Sateia, M. (1993) Measuring quality of care in psychiatric emergencies: construction and evaluation of a Bayesian index. *Health Services Research, 28*, 131–158.

Gustafson, D. H., Sainfort, F. C.,Van Konigsveld, R., & Zimmerman, D. R. (1990). The Quality Assessment Index (QAI) for measuring nursing home quality. *Health Services Research, 25*, 97–127.

Gustafson, D. H., Tianen, B., & Greist, J. (1981). A computer-based system for identifying suicide attempters. *Computers and Biomedical Research, 14*, 144–157.

Hall, G. (1992). The local educational change process and policy implementation. *Journal of Research in Science Teaching, 29*(8), 877–904.

Hall, G., & Hord, S. M. (1984). *Change in schools: Facilitating the process.* New York, NY: State University of New York Press.

Hall, G., & Loucks, S. (1977). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal, 14*(3), 263–276.

Hernandez, S. R., & Kaluzny, A. D. (1994). Organizational innovation and change. In S. M. Shortell & A. D. Kaluzny (Eds.). *Health care management: Organization, design, and behavior* (3rd edn, pp. 294–315). Albany, NY: Delmar.

Huberman, A., & Miles, M. (1984). *Innovation up close: How school improvement works.* New York: Plenum.

Knapp, M. S. (1995). How shall we study comprehensive, collaborative services for children and families? *Educational Researcher, 24*(4), 5–16.

Levenson-Gingiss, P., & Hamilton, R. (1989). Determinants of teachers' plans to continue teaching a sexuality education course. *Family and Community Health, 12*(3), 40–53.

Marks, E., L. & Marzke, C. H. (1994). *Health caring: A process of evaluation of the Robert Wood Johnson Foundation's school-based adolescent health care program.* Princeton, NJ: Mathtech, Inc.

McLaughlin, M. W. (1990). The Rand change agent study revisited: Macro perspectives and micro realities. *Educational Researcher, 19*(9), 11–16.

Ottoson, J. M., & Green, L. W. (1987). Reconciling concept and context: Theory of implementation. *Advances in Health Education and Promotion, 2*, 353–382.

Parcel, G. S., Taylor, W. C., Brink, S. G., Gottlieb, N. H., Engquist, K., O'Hara, N. M., & Erikson, M. P. (1989). Translating theory into practice: Intervention strategies for the diffusion of a health promotion intervention. *Journal of Family and Community Health, 12*(3), 1–13.

Roberts-Gray, C. (1985). Managing the implementation of innovations. *Evaluation and Program Planning, 8*, 261–269.

Roberts-Gray, C., & Scheirer, M. (1988, Winter). Checking the congruence between a program and its organizational environment. In K. J. Conrad & C. Roberts-Gray (Eds.). *New Directions for Program Evaluation: No. 40, Evaluating program environments.* (pp. 63–82). San Francisco: Jossey-Bass.

Rogers, E. M. (1983). *Diffusion of innovations.* (3rd edn). New York: Free Press.

Rogers, E. M., & Shoemaker, F. F. (1971). *Communication of innovations: a cross cultural approach* (2nd edn). New York: Free Press.

Rothman, J., Erlich, J. L., & Teresa, J. G. (1981). *Changing organizations and community programs.* Beverly Hills: Sage Publications.

Rowse, G., Gustafson, D. H., & Ludke, R. (1974). Comparison of rules for aggregating subjective likelihood ratios. *Organizational Behavior and Human Performance*, *12*, 274–285.

Scheirer, M. (1990). Life-cycle of an innovation: Adoption vs discontinuation of the fluoride mouth-rinse program in schools. *Journal of Health and Social Behavior*, *31*(2), 203–215.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge; New York: Cambridge University Press.

Yin, R. K. (1979). *Changing urban bureaucracies: How new practices become routinized*. Lexington, MA: Lexington Books.