

Planning and Conducting **FORMATIVE EVALUATIONS**

**Improving the Quality of Education
and Training**

Martin Tessmer

**Planning and
Conducting
FORMATIVE
EVALUATIONS**

Planning and Conducting FORMATIVE EVALUATIONS

**Improving the Quality of Education
and Training**

Martin Tessmer



KOGAN PAGE

London • Philadelphia

First published in 1993
Reprinted 1995, 1997

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers at the undermentioned address:

Kogan Page Limited
120 Pentonville Road
London N1 9JN

© Martin Tessmer, 1993

British Library Cataloguing in Publication Data

A CIP record for this book is available from the British Library.

ISBN 0 7494 0801 4

Typeset by Saxon Graphics Ltd, Derby
Printed and bound in Great Britain by
Biddles Ltd, Guildford and King's Lynn

Contents

Glossary	7
Introduction: A Note to My Readers About This Book	8
Why this book was written	8
Who this book is for	8
What's new in this book	9
How this book is organized	9
Using this book	10
How this book was formatively evaluated	10
Chapter One: Central Questions and Issues in Formative Evaluation	11
What is formative evaluation?	11
When did formative evaluation originate?	11
Why do formative evaluation?	13
What instruction is evaluated?	14
What types of formative evaluation are there?	15
When is formative evaluation used in a project?	17
Who participates in the evaluation?	18
What do you learn from an evaluation?	19
When is formative evaluation most necessary?	20
How do you conduct a formative evaluation?	21
What tools are used?	21
What are the problems in conducting an evaluation?	22
Summary	24
Chapter Two: Planning the Evaluation	25
Introduction	25
Some basic planning questions	26
The evaluation planning process	38
Summary	43
Example of formative evaluation planning	43
Appendix	45

Chapter Three: Expert Review	47
Introduction	47
Planning considerations	49
Conducting an expert review	55
Summary	67
Example of expert review	67
Chapter Four: One-to-One Evaluation	70
Introduction	70
Planning considerations	73
Conducting a one-to-one evaluation	80
Summary	97
Example of one-to-one evaluation	98
Chapter Five: Small Group Evaluation	101
Introduction	101
Planning a small group evaluation	104
Conducting a small group evaluation	111
Summary	134
Example of small group evaluation	135
Chapter Six: Field Test Evaluation	137
Introduction	137
Planning a field test	139
Conducting a field test	142
Summary	152
Example of a field test	152
References	154
Index	158

Glossary

CD-ROM refers to Compact Disc, Read Only Memory, a media format where the instruction is stored on the compact disc medium.

CBT is the acronym for Computer Based Training, and can include videodisc, hypertext/media, and CD-ROM formats for tutorials, simulations, drills, etc.

Constructivism is a learning theory that knowledge is constructed, and not discovered as an objective reality. Instruction should be designed to facilitate learners' personal knowledge construction.

Distance learning includes all forms of instruction where students and teacher are spatially separate: two-way video, radio, correspondence courses.

Hypertext is a form of organizing instruction in a nonlinear format that allows learners to reference information based on their individual needs.

Hypermedia is instruction organized in the same format as hypertext, but with a wider variety of media; sound, animation, text, video, etc.

Task analysis is that instructional design stage where designers analyse instruction to discover the structure or sequence of what needs to be learned.

The **instructor** need not be a teacher or trainer. It refers to anyone who may administer or coordinate learning: lab manager, discussion coordinator, tutor.

Introduction

A Note to my Readers about this Book

Why this book was written

'Formative evaluation' is the systematic tryout of instruction for purposes of revising it. The method is a valuable and respected part of instructional design. However, as a practitioner of formative evaluation, I have never found a book that I could use as a reference work for the subject. Such a book would have guidelines, job aids and samples for each type of formative evaluation. The practitioner would be able to take such a book off her shelf, consult the needed sections, and quickly brush up on the whys and wherefores of specific types of formative evaluation.

As a teacher of formative evaluation courses, I could not find a satisfactory textbook written from the perspective of instructional design and for instructional designers. Such a textbook would offer both theory and procedures of formative evaluation. It would help students learn to plan and conduct evaluation in the context of the real-world exigencies of limited time and resources, other instructional design tasks, and recalcitrant subjects. Such a book would help the reader become versed in the general history and concepts of formative evaluation, but would also help the novice evaluator conduct successful evaluations.

Who this book is for

This book is designed for anyone who may be involved in developing instruction. This includes students and professionals, in both education and training. Even the experienced formative evaluator should find something new in each chapter.

Wherever possible I have tried to include terms and examples that are relevant to both educators and trainers. Thus, the words 'instruction' and 'instructor' are used whenever possible, to denote both education and training, teachers and trainers. In some places the word 'administrator' is used to designate anyone who administers the instruction: instructors, lab managers, graduate assistants, librarians. Other relevant terms are defined in the Glossary.

Above all, this is a book about how to do formative evaluations. This is not a book about the history of formative evaluation (although there is some of that) nor is it a statistics-laden book on how to conduct experimental or quasi-experimental group evaluations. A professional evaluator might be disappointed with the lack of discussion of group statistical measures that can be used to measure statistically

significant learning gains. However, in my experience the great majority of evaluations do not have the luxury of sufficient subjects for measures of statistical significance. Those who do are usually well-funded enough to hire a specialist to do this anyway! Therefore, this book concentrates on the most frequent evaluation scenarios that you will encounter, those where you have one or two experts or a couple of students, or (for small group or field tests) 8-15 students.

What's new in this book

In my opinion you will find several emphases in this book that you would not find in any formative evaluation text that I have encountered. This book has a strong emphasis on *planning* the overall stages and sequence of the evaluation, and Chapter Two is entirely devoted to it. The book details the art of asking the right *questions* in the evaluation – a critical and tricky issue that needs more treatment than it has received in formative evaluation literature. The book also relates formative evaluation to all types of media, not just text or video. This includes new technologies such as hypertext and multimedia as well as overlooked ones such as workshops and lectures. Because formative evaluation can be an intensely interpersonal encounter between subject and evaluator, *interpersonal issues* and problems are discussed. The way to consider and make *revisions* is part of every chapter, and is a critical and difficult issue that has received little coverage elsewhere. Finally, the book outlines some *alternative forms of formative evaluation*, ones that do not fit the classic types but are nonetheless useful. These include self-evaluation, two-on-one evaluation, think-aloud protocols, continuous assessment and rapid prototyping.

How this book is organized

Chapter One explains the concept of formative evaluation; its origin, essential characteristics, value, and limitations. The second chapter outlines the general procedure for planning an evaluation; how to start thinking about what and who to evaluate, the types of information to look for, and the way to fit formative evaluation into even the most resource-limited project. The next four chapters treat the classic stages of formative evaluation: expert review, one-to-one, small group and field test. Each chapter covers a specific evaluation stage and details what that stage is, how to plan it, how to do it, and how to cope with special problems that may arise.

Each chapter contains one or two 'sidebars' of important information enclosed in a text box. These sidebars deal with important formative evaluation issues and methods, but are not part of the regular 'flow' of the text chapter. My student reviewers have read the sidebars as they read the chapter (instead of after) and did not believe that they detracted from the reading of the regular text.

Using this book

As you may have guessed, this book is designed to be both a text and reference work. In my view, the best way to use this book is to read through the entire book as a text, and to reference later chapters as needed for your particular evaluation needs. The chapters on expert, one-to-one, small group and field test evaluations are designed to be independent of each other. That is, each chapter has its own procedures, questions, and issues about that particular type of formative evaluation.

Since each chapter is a self-contained instructional unit, you may individually reference chapters to apply to your real-world evaluation needs. For example, you may find that three months after reading this book you are commissioned to conduct a small group evaluation. You could then take this book and reference the chapters on planning an evaluation and small group evaluation (it's a good idea to always review the planning chapter, even if you know what evaluation stages you will use). In using only these two chapters, you can review the specific concepts and skills you need for your evaluation situation. However, I would recommend reading the entire book first, since even experienced designers may find a few new ideas and 'tricks of the trade!'

How this book was formatively evaluated

This book has undergone an expert review and a field test evaluation. The expert review was conducted via mail, phone and personal interview with Professor Walter Dick of Florida State University, one of the world's foremost figures on formative evaluation and instructional design. Professor Dick read each chapter and noted his comments, as well as answering a prepared set of questions relating to each chapter. The book was also used as an instructional text in my Summer 1992 graduate class on formative evaluation. The students wrote comments on each chapter of the text, answered a questionnaire about each chapter, and furnished class comments on a regular basis. The comments by all participants were insightful and invaluable, and I am grateful to them all for their help.

Student reviewers		
Joanna Dunlap	Leann Fields	Leslie Harris
Juliana McCahan	Brad Michels	Barbara Youmans
Thanks to you all!		

Central Questions and Issues in Formative Evaluation

What is formative evaluation?

Perhaps the best way to begin is to define each word in the term. 'Formative' is used in a developmental sense, as children are in their 'formative' or developing years and are susceptible to growth and change. The evaluation target is instruction in its formative stages, instruction that is developing and not yet finished or 'grown up' and is thus amenable to revision.

'Evaluation' is a data gathering process to determine the worth or value of the instruction, of its strengths and weaknesses. The identified strengths and weaknesses are used to revise the instruction to improve its effectiveness and appeal. Thus, 'formative evaluation' is a judgement of the strengths and weaknesses of instruction in its developing stages, for purposes of revising the instruction to improve its effectiveness and appeal. The evaluation is conducted by collecting data about the instruction from a variety of sources, using a variety of data gathering methods and tools.

When did formative evaluation originate?

It is always difficult to track the origin of an idea or method, but we do know that instructional materials were tried out and tested for improvement during the audiovisual movement of the 1920s when educational film was evaluated (Williams, 1983; Cambre, 1981). The use of formative evaluation (albeit under different names) was used through the 1930s to the '50s in various projects, many of them upon educational film. Be that as it may, formative evaluation did not become a systematic design method until the late 1960s, when formal models of the process were drawn up, and tryout-and-revision became an integral part of the

programmed instruction movement. Scriven (1967) then attached the name 'formative evaluation' to a revisionary process that had been referred to as 'tryout', 'developmental testing', and other terms. Today, there are still a number of terms used in place of formative evaluation, terms such as 'pilot test', 'formative assessment', 'dry run' 'alpha/beta testing', 'quality control' (my favourite) and the impressive 'learner verification and revision!'

Until the 1960s evaluation of instruction was usually *summative*. Summative evaluation was of materials in their 'adult' (not formative) stage, meaning that they were already completed and in final form. The purpose of the evaluation was often comparative, to determine if the instruction was better than some other form of instruction. In terms of the design and improvement of instruction, this method was similar to locking the stable door after the horse had bolted, since the materials were already completed and disseminated, and could only be revised if a new version was designed. Markle (1989) refers to this as the difference between using evaluation to *prove* (summative) versus *improve* (formative), and Baker and Alkin (1973) as the difference between evaluation for validation (summative) versus evaluation for revision (formative).

With the advent of large scale curriculum projects in the 1960s, designers began to more strongly heed the words of Ralph Tyler (1942) and to see curriculum development as a continuous process of tryout and development for improvement. Thus, designers including Ken Komoski, Eva Baker, George Geis and Walter Dick began to set forth principles and procedures for formatively evaluating instruction. Formative evaluation then grew (with many growing pains) to become a recognized and respected part of systematic instructional development. Braden (1987) believes that formative evaluation distinguishes instructional design from all other instructional improvement methods.

Formative evaluation is not an activity to prove or validate the effectiveness of your instructional design, it is part of the instructional design itself. In other words, this is not an add-on process to find out if your instruction is effective, it is a problem-finding part of a design and product development process. The importance of this distinction is that many organizations still do not perceive formative evaluation as necessary to instructional development. They say, 'If you have good designers and good producers, why have formative evaluation?' They see formative evaluation as a waste of resources, or as a sign of insecurity about quality of the instructional development effort. However, many training and education organizations that utilize formative evaluation do so because they understand it is part of the design of effective instruction, and because they see it as cost-savings measure to economically 'debug' instruction and increase client satisfaction. This perspective is one that designers often must foster within the organizations in which they work.

Why do formative evaluation?

Over the last 30 years, a number of empirical studies have shown that formatively evaluating instructional materials has resulted in revised instruction that produces statistically significant increases in student performance over the original, unevaluated versions of the instruction (Nathenson and Henderson, 1980). These improvements have been reported on all types of instruction: computer-based instruction, simulations and games, text, and multimedia. Beneficial results have also been reported upon every evaluation stage: expert review, one-to-one evaluation, small group and field test. Even the use of a single learner for an evaluation has resulted in improved materials (Lowe, Thurston and Brown, 1983). Thus, there is evidence that using formative evaluation can improve the learning effectiveness of instructional materials.

Can formative evaluation make instruction more motivating or interesting as well as more effective? There is not enough evidence to say. Historically, most formative evaluation studies have measured student performance gains and ignored measures of student attitude or acceptance (or instructor attitude/acceptance!). However, evaluators have used both experts and students to evaluate the interest and acceptability of instruction to its users, and have obtained suggestions on how to revise instruction to meet these goals (Flagg, 1990; Nathenson and Henderson, 1980). Certainly, formative evaluation can be used to obtain criticisms and suggestions on the interest/motivation of the instruction to its users.

Do instructional designers actually utilize formative evaluation in their design projects? Zemke (1985) surveyed readers of *Training* magazine and found that more than 60 per cent of the respondents used some form of formative evaluation in their projects, more than most instructional design activities listed in the survey. Tessmer and Wedman (1992) in a later survey of corporate training professionals, found that almost half their respondents used some type of formative evaluation in their projects. Formative evaluation then appears to be part of the 'real world' of instructional design.

Even though formative evaluation is frequently used by practitioners, this does not mean that it is readily accepted by all organizations. Many managers or clients do not understand the purpose or utility of formative evaluation. They may think that:

- evaluation is something that you do after the product is finished,
- all these evaluations are unnecessary if you have competent and experienced designers,
- their personal evaluations are sufficient.

This means that instructional designers must often argue for including formative evaluation in a project, and must demonstrate its benefits. Designers argue for the evaluation as a quality control measure of

instructional products. They demonstrate its benefits by citing the errors identified and revisions made in other formative evaluation projects, and explaining the cost savings that can result from revising instruction before it is finalized.

What instruction is evaluated?

The most popular target for formative evaluation is instructional materials. Surveying formative research for the last 30 years, it becomes obvious that text is the most evaluated type of instructional material, with numerous evaluations also undertaken with film, video, multimedia, and computer-based instruction. In fact, any form of material can be competently and profitably evaluated, including CD-ROM, interactive video, hypertext/media, and broadcast television or radio. The evaluation questions may change depending on the media attributes of the format evaluated, but the overall formative evaluation process remains the same.

Many designers somehow think that instructional materials are the sole target of evaluation. They ignore live instruction as evaluation candidates. In fact, any form of instruction that can still be revised before being used in the 'real world' can be formatively evaluated (Baker and Alkin, 1973). Workshops, seminars and lectures are all viable formative evaluation candidates. Person-led instruction is really just another form of media, since the person acts as an instructional delivery system with a unique set of media attributes. As a 'media', person-led instruction will also have its own type of evaluation questions that are pertinent to that type of media. As with other media, the overall process of formative evaluation is still applicable.

Formative evaluation is usually applied to materials being developed by the design team, but finished materials can also be evaluated for their adaptability to the instructional environments in which they will be used. As long as the purpose of the evaluation is to 'revise' the instruction by reorganizing or supplementing it, the evaluation can be a type of formative evaluation. For example, instructors can evaluate commercially prepared materials to see if they fit in with their learners, classroom schedules, curriculum objectives, and physical facilities. The results of the evaluation may mean that only parts of the instruction are used, or that new materials are developed to supplement the inadequacies of the commercial ones. Similarly, learners can evaluate the instruction's learnability and appeal to determine if adaptations must be made to it.

It is important to remember that any size of instruction can be formatively evaluated. The target for a formative evaluation can be a unit, lesson, course or curriculum. Frequently, the target for evaluation is a unit or lesson simply because of the considerable time and money required to evaluate whole courses or curricula. As we shall see in the next

chapter, designers must often choose the amount and type of formative evaluation they will do based on the constraints of the project. Where there is 'too much' instruction to evaluate within project constraints, such as a course or curriculum, they may select smaller segments to more thoroughly evaluate, segments that may allow them to generalize their findings and revisions to unevaluated segments.

The focus of this book is on the formative evaluation of instructional *products* such as texts, lectures, and multimedia instruction. The formative evaluation of instructional *programmes* is also an important part of education and training, but is conducted less frequently by instructional designers than formative product evaluation. Programmes could be an instructional service such as a writing programme, or a curriculum-wide project such as a maths literacy programme. Programme evaluation is frequently conducted by educational researchers, and can be a complex and time-consuming process. Readers who are interested in formative programme evaluation are referred to Bloom *et al* (1971).

What types of formative evaluation are there?

There are four classically recognized types of formative evaluation, which have these general characteristics:

- expert review – experts review the instruction with or without the evaluator present. The experts can be content experts, technical experts, designers or instructors.
- one-to-one evaluation – one learner at a time reviews the instruction with the evaluator and comments upon it.
- small group – the evaluator tries out the instruction with a group of learners and records their performance and comments.
- field test – the evaluator observes the instruction being tried out in a realistic situation with a group of learners.

Ideally, the evaluator would conduct expert and one-to-one evaluations first, revise the instruction, and then conduct a small group evaluation and 'polish' the instruction once again (see Figure 1.1). Revisions may be made between the one-to-one and expert reviews, or the evaluations may be carried out simultaneously. The final step is then to field test the instruction in the learning environments for which it was intended.

Within these four types there are many variations. The number and variety of experts in an expert evaluation can vary, just as the number and variety of learners in a one-to-one can vary. The reality of the situation also varies, so that a small group evaluation is done in learning situations that are more realistic than one-to-ones, with field tests most closely emulating real-world situations and procedures. In a one-to-one evaluation the evaluator works closely with the learners, while in a field test the evaluator may be present only as an observer.

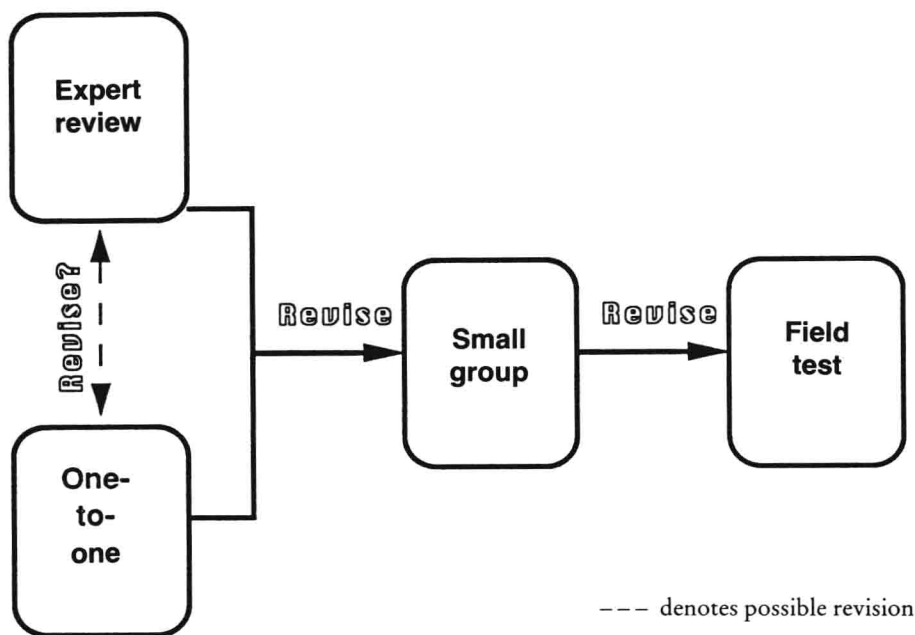


Figure 1.1 General sequence of formative evaluation types

The methods also vary in the time and effort to conduct them, and the degree of sophistication of the evaluated materials. A single one-to-one or expert evaluation may require little time or money to conduct. Often these two evaluation stages are used with instruction in its rough stages: scripts of videos, photocopied texts with hand-drawn illustrations, outlines of workshop content and activities. Small group and field test instruction may be more resource-intensive and be used with more polished instruction: the video is produced, the text is in publishable form, and the workshop is conducted as a dress rehearsal. Because of these differences, expert reviews and one-to-ones are often completed first, to eliminate mistakes and inaccuracies before materials are produced in a more sophisticated format that costs more to revise.

In addition to these four formative evaluation stages, there are variations that can be used during an evaluation, such as:

- self-evaluation – designers or the design team evaluate their own instruction.
- expert panels – teams of experts discuss the instruction together with the evaluator.
- two-on-one evaluation – two learners review the instruction with the evaluator.