CHAPTER II

REVIEW OF THE LITERATURE

Significant Historical Factors

According to Keegan (2000), teaching at a distance began 175 years ago in large part due to the development of postal communications and rail transport, the technologies associated with the Industrial Revolution in Northern Europe and North America. These emerging technologies led to the development of correspondence studies, which began to appear in the 1830’s (Hanson, Maushak, Schlosser, Anderson, Sorensen, & Simonson, 1996). Chautauqua College of New York offered degrees via correspondence study from 1883 to 1891. During this same time period, the University of Chicago created the largest correspondence program in the United States. The Chicago program was established to offer courses to learners who could not afford to participate socially or financially at full-time learning institutions (McIsaac & Gunawardena, 1996).

Early correspondence programs were not highly regarded. As delivery systems evolved, the effectiveness of distance education increased (McIsaac & Gunawardena, 1996). Traditionally, the primary communications technology utilized in distance education was print-based text. Distance education used other forms of media: radio during World War I, and later television in the 1950’s (McIsaac & Gunawardena, 1996). Distance education was transformed with the establishment of the University of South Africa in 1962 and the founding of the British Open University in 1969 (Simonson, Smaldino, Albright, & Zvacek, 2000; McIsaac & Gunawardena, 1996). These full degree programs introduced the use of technologies to supplement print-based instruction in delivery and presentation.
The telecommunications electronics revolution in the 1980’s hastened the creation of faster computer chips as well as broadband technologies. Broadband technologies provided the means to deliver images, compressed video, and multimedia interactions to individuals and institutions (Keegan, 2000). Moreover, the costs of technology continued to decline (Dede, 1990). In the early 1990’s, the first graphic-based Web-browser sparked the inception of the World Wide Web (WWW) and the development of hypertext markup language (HTML).

Since the 1980s, advances in information technology have increased at an astonishing rate. Use of personal computers and use of the web have become commonplace in schools, libraries, and military training facilities, as well as in the home.

The Internet is a system of linked computers tied together by a common communication language known as Internet Protocol (IP). Internet Protocol allows the exchange of “packets” of information between specified computer addresses (Brooks et al., 2001).

Historically, “time” and “place” distinguish distance education from traditional, face-to-face education. Distance education can be defined as an educational process of formal or informal instruction using print or electronic communications media in which instructors and learners are separated by time and/or geographic location (McIsaac & Gunawardena, 1996; Yates & Tilson, 2000). Distance education literature cites various other characteristics to complete the definition. For example, Keegan (1986) defines distance education as a planned and systematic activity. Keegan (2000) indicates that the essential part of distance education is the separation of the instructor from the learner,
and of the learner from the learning group. Keegan (1988) provides several defining characteristics:

1. The quasi-permanent separation of student and instructor distinguishes distance education from conventional face-to-face education.

2. An educational institution is involved with the administration and organization of the learning content and support services.

3. The instructor and learners are united through the use of technology and media to deliver the course content.

4. Technology is used to communicate so that students may initiate and benefit from dialog.

5. Distance education also features the quasi-permanent absence of learning groups, along with content designed for individual learning.

As technology has evolved, so has the definition of distance education. The communications revolution of the 1980’s created opportunities to teach and learn face-to-face-at-a-distance (Hanson et al., 1996; Keegan, 2000). Broadband technologies and advancements in computer simulations and tutorial programs will force the traditional definition of distance education to evolve. The concept of distance itself may evolve as synchronous communications become available (Hanson et al., 1996). With tutorial and simulation programs, the instructor may not be required as an instrument of teaching. Students interacting with technology designed by subject matter experts and instructional designers in a systematic way might affect the need for instructor and student interactions, as well as for peer-to-peer interactions.

Adult Distance Education and Distance Training
It is important to define the population a researcher is examining. Tight (1996) defines an adult as an individual who has reached a certain status in life, capable of providing for himself or herself, and responsible for making individual choices. According to Tight, it is not how we see ourselves but how others view us. More precisely, “adulthood is measured as an ethical status resting on the presumption of various moral and personal qualities” (Paterson, 1979).

In contrast to distance education, distance training involves content that is more specific, featuring decreased breadth and shallower knowledge. Training has compressed time commitment in comparison to distance education (Tight, 1996). Keegan (2000) indicates that training is usually created for individual-based systems. Another distinction between education and training is the desired outcome of the activity. Distance education provides long-term benefits to the individual, leading to a degree or professional certification. Training, on the other hand, provides near-immediate benefits to the individual’s job skills. Training is usually related to a learner’s career (Keegan, 2000).

Research in distance education has been primarily focused on education rather than training. According to Wisher, Champagne, Pawluk, Eaton, Thornton, & Curnow (1999), only 32 of the 241 studies cited by Russell (1998) focused on adult training.

**Early Childhood Education**

In 2002, there were 2.3 million paid U.S. providers of primary care to infants and toddlers in a given week (Burton, Whitebook, Yung, Bellm, Wayne, Brandon, & Maher, 2002). The number of women in the workplace has risen dramatically. In 1996, the number of working mothers with children under the age of six tripled, climbing from 20
percent in 1960 to 62.3 percent in 1996. Roughly half of working mothers return to their jobs within the first year of their infants’ life (Organization for Economic Cooperation and Development Country Note, 2000). As a result, the percentage of children in care centers also is increasing.

Early childhood programs vary greatly in quality. Most states do not require any specialized training to handle infants and toddlers. In a survey of 400 daycares, 40% were found to have dangerous physical environments that jeopardized children’s health and safety. Only 9% of daycares were found to be “good” (Organization for Economic Cooperation and Development Country Note, 2000). According to Love (1997), only 8% of infant classrooms and 24% of preschool classrooms were good or excellent in quality. Quality of the environment has been correlated to outcomes (Love, 1997; Peisner-Feinberg, Burchinal, Clifford, Yazejian, Culkin, Zelazo, Howes, Byler, Kagan, & Rustici, 1999). In addition, lower quality environments suffer from decreased levels of engagement, a critical factor in learning (McWilliam, Maxwell, Ridley, de Kruif, Raspa, Appanaitis, & Harville, 2001).

Comparing childcare in Nebraska to national averages, research shows an estimated 37% of programs were categorized as “good” in quality, 49% were rated as “mediocre,” and 14% were rated as “poor” (Edwards, Knoche, Raikes, Raikes, Torquati, & Wilcox, 2002).

Children in low quality environments are at a developmental risk that can lead to less competent and productive members of society later in life (Peisner-Feinberg et al., 1999; Szanton, 1997). Children in high quality childcare had better language and math skills than children who attended low quality programs (Peisner-Feinberg et al., 1999).
Some children have a greater risk of lower math abilities and social maladjustment when they experience low quality of care (Peisner-Feinberg et al., 1999). The effects of inadequate childcare programs are lasting and have been observed through the second grade, four years beyond the childcare experience (Peisner-Feinberg et al., 1999).

In 2000, the average hourly wage was $7.86 per hour for childcare workers and $9.66 per hour for preschool teachers in the United States (Laverty, Siepak, Burton, Whitebook, & Bellm, 2002). By comparison, the average hourly wage for a secretarial position was $11.98 per hour. According to Laverty et al. (2002), the average per hour wage in Nebraska was $7.12 for childcare workers and $9.10 for preschool teachers. In Nebraska, a childcare worker receives less per hour than the average wage of a parking lot attendant (Laverty et al., 2002). In Nebraska, 77.5% of infant and toddler caregivers make less than $15,000 a year (Raikes, 1998). The average annual salary for childcare providers in 2002 was estimated at $14,700, below the poverty line for many providers (Edwards et al., 2002).

Most providers receive benefits such as paid vacation days (93.1%) and sick leave (73%) in Nebraska. Other benefits include paid days to attend professional meetings (67.4%) and reduced or no tuition for their children (62.4%). A little more than half (51.4%) of providers receive health insurance for themselves, and 41.6% receive health insurance for their family. One third (33.3%) of providers receive retirement benefits (Edwards et al., 2002).

Burton et al. (2002) provide an insight into the demographics of childcare providers. Thirty-three percent of providers have a bachelors degree, 47% have some
college, and 20% have a high school diploma or less (Burton et al., 2002). The turnover rate for childcare providers is between 19% and 50% (Burton et al., 2002; Love, 1997).

Training and continuing education for providers has been correlated to higher quality early childhood education (Edwards et al., 2002; Peisner-Feinberg et al., 1999). Fewer than 30% of providers were involved in continuing education or training in their specialty area, however (Raikes & Jones, 1998).

The demographics of providers in the state of Nebraska are similar to those of other midwestern states. The average childcare provider in Nebraska is 38 years old and is married. Most providers have an average of over five years experience. An average of 14% of providers have obtained a bachelors degree or higher. An additional 18% have a two-year degree. Thirty-two percent of providers have some college education, while 28% have a high school diploma (Edwards et al., 2002).

There are few training and continuing educational opportunities for providers in the area of early childhood education. Three main barriers to training exist for care providers: lack of time to participate; significant costs associated with attending training; and no clear motivation to spend the time and money to attend the training (Raikes, 1998). In general, centers have problems finding and paying for adequate training programs. As a result, most training is sporadic, duplicated, and fails to meet the needs of the care providers (Peisner-Feinberg et al., 1999).

**Effectiveness of Educational Technology**

Historically, technology in distance education has been used as a communication and delivery medium (Keegan, 2000). However, as technology has evolved, so to has its use. Distance education can be centered on the technology, with the input of the
instructor used to convey information (Belanger & Jordan, 2000). Pacing and feedback are incorporated into the instructional design and carried out by the media (Simonson et al., 2000). As a consequence of the heavy reliance on technology, research has attempted to identify the effectiveness and efficiency of technology in distance education (Keegan, 2000).

Meta-analyses of research on computerized multimedia suggests an effect size of approximately 0.30 to 0.38 standard deviations on assessment scores (ADL, 2001; Bayraktar, 2002; Fletcher-Flinn & Gravatt, 1995; Kahalili, 1994; Kulik & Kulik, 1991; Kulik, 1994; Liao, 1992; Melmed, 1995). Typically, the effect size is smaller for programs lasting over four weeks (Kahalili, 1994; Liao, 1992). Cavanaugh (1999) reports an effect size of 0.147 for K-12 distance education courses (Cavanaugh, 1999). Similarly, Christmann and Badgett (2000) cite an effect size of 0.187 to 0.209 for secondary students (Christmann & Badgett, 2000; Christmann, Badgett, & Lucking, 1997). They report an effect size of 0.127 in college students (Christmann & Badgett, 2000). With the addition of multimedia, the effect size can be raised to 0.50 (ADL, 2001; Fletcher, 2001).

In a study of over 10,000 subjects in over 30 experiments, Vinsonhaler and Bass (1972) found that drill and practice computer assisted instruction (CAI) significantly increased student's achievement in mathematics, language arts and basic skills tests. In elementary schools where CAI was used, students moved from the 50th percentile to the 66th percentile in mathematical abilities. Hartley (1977) found passing rates of California
community college students taking remedial math classes increased from 15% to 25% (Baker, Hale, & Gifford, 1997).

Studies also indicate that the time for instruction is compressed when technology is used (Kulik & Kulik, 1991). Computer based instruction (CBI) took approximately two-thirds of the time required for traditional courses in 32 post-secondary classrooms (Kulik & Kulik, 1991). Adult learning courses were compressed by an average of 24% (Kulik, 1994). Schmeekle (2000) found that the amount of instructional time was reduced by 50% for adult distributed training in law enforcement. In a more generalized view, Fletcher (2001) states that the average instructional time is reduced by 30 percent with computer-based training. Meta-analysis also found effect sizes across all age groups. Adult studies had an average effect size across all areas of 0.25 (Kulik & Kulik, 1991).

According to Clark (1994b), meta-analytic research in computer-based training may have fundamental flaws. Meta-analytic studies may not have accounted for the confounding of the media variables with instructional method (Clark, 1994b). Clark (1994a) indicates that it was the instructional method used and not the media that produced the positive outcomes in meta-analytic studies. Additionally, Clark states that the results of these effect sizes were not significant (Carter, 1996).

Moreover, there are many critics of research methodologies in education technology. Todd Oppenheimer (1997) succinctly summarizes problems with media-comparison studies:
One meta-analysis (a study that reviews other studies -- in this case 130 of them) reported that computers had improved performance in "a wide range of subjects . . .”

Unfortunately, many of these studies are more anecdotal than conclusive. Some, including a giant, oft-cited meta-analysis of 254 studies, lack the necessary scientific controls to make solid conclusions possible. The circumstances are artificial and not easily repeated, results aren't statistically reliable, or, most frequently, the studies did not control for other influences, such as differences between teaching methods. . . . "The research is set up in a way to find benefits that aren't really there," Edward Miller, a former editor of the Harvard Education Letter, says. "Most knowledgeable people agree that most of the research isn't valid. It's so flawed it shouldn't even be called research. Essentially, it's just worthless."

**Media Comparison Studies**

Educational researchers have examined the influence of media on learning (Lockee, Burton, & Cross, 1999). Several studies compare the results (usually learning outcomes) of instruction using media to the results of the traditional classroom. These are defined as media comparison studies. In 1983, Richard Clark ignited a debate on the relevance of such media studies. Clark begins the debate by stating that, “Media do not influence learning under any conditions.” He states that media are “mere vehicles that deliver instruction but do not influence student achievement anymore than the truck that delivers out groceries causes changes in our nutrition” (1983). Clark updated his analogy in 1994, comparing the various methods of pharmaceutical delivery methods (pills,
suppositories, IV, and injections) with their effectiveness. Clark indicates that the delivery method does not increase the patients’ health; rather, improvement is the result of the active chemical ingredients (1994a). Clark believes that media cannot be separated from instructional design, and do not influence learning improvement (Tennyson, 1994). Moreover, according to Clark, media alone does not affect learning. Media can reduce the costs and increase the efficiency of learning, but the use of an adequate teaching method alone will ultimately influence learning (Clark, 1994b). In well-designed media comparison studies, Clark suggests that no substantial differences in learning outcomes should be expected.

Furthermore, Clark (1994a) indicates that there is no single media attribute that cannot be replicated by similar media attributes. This is central to Clark’s argument. He terms this the replaceability test: “If a treatment can be replaced by another treatment with similar results, the cause of the results is in some shared (and uncontrolled) properties of both treatments” (1994). Clark cites the literature collected by Russell to bolster his argument that, over the past 70 years of comparison study research, no significant differences between mediated and traditional treatments exist (Russell, 2001).

There are alternative views regarding the influence of media on learning. Jonassen, Campbell, & Davidson (1994) see media as a way to empower the learner; they conclude media can influence learning. Kozma (1994) and Reiser (1994) also contend that media influence learning. According to Kozma (1994), media can be treated as an independent variable that can be shown to improve learning over traditional teaching methods. Likewise, Reiser (1994) indicates that media have attributes that can influence learning in certain situations.
Kozma (1994) believes media and method are integrated and connected within the environment and learning situation. Kozma concedes that past research does not indicate a relationship between media and learning. However, he believes a relationship will exist between media and learning. Kozma (1994) indicates that learning occurs because of a unique mix of methods, technologies, and initiatives taken by the learner in a learning environment. Due to the relationship between media and learning, the research questions should ask, “In what ways can we use the capabilities of media to influence learning for particular students, tasks, and situations?” (1994).

It is plausible that greater research controls in comparison research may provide increased levels of understanding. A comparison research model may provide the necessary checks for validity in instruments and procedures. Moreover, reporting of sample size, power estimates, and effect size may increase the reliability of research findings. Additional reporting of the population under study and a description of the content and context of learning may lead to better understanding of the effectiveness of unique combinations of media and method on a given population for a given objective.

**Research Methods**

The following validity types are reported in the literature concerning experimental and quasi-experimental research: a) statistical conclusion validity, b) internal validity, c) construct validity, d) external validity, and e) face validity (Borg, Gall, & Gall, 1993; Creswell, 2002; Shadish, Cook, & Campbell, 2002).

Shandish et al. (2002) define statistical conclusion validity as “the validity of inferences about the correlation (covariation) between treatment and outcome.” Threats to statistical conclusion validity include: a) low power, b) unreliability of measures, c)
unreliability of treatment implementation, d) heterogeneity of units, and e) inaccurate effect size estimation.

Threats to internal validity confound the observed relationship between the treatment and the outcome, due to problems with experimental procedures or participant experiences (Creswell, 2002; Shadish et al., 2002). Several writers, (Campbell & Stanley, 1963; Creswell, 2002; and Shadish et al. 2002), discuss factors that threaten internal validity:

1. *Ambiguous temporal precedence* indicates problems with the order of variables.
2. The *history variable* is comprised of events that take place between the first and second measurement.
3. *Maturation* is the process of time on participants of the study.
4. *Testing* is the effect of a first test on the outcomes of a second test.
5. *Instrumentation* includes changes in the calibration of an instrument or an observer.
6. *Statistical regression* can be seen when extreme scores are the basis of selection for a group.
7. *Biases* in the selection of comparison groups can be a factor.
8. *Attrition* (experimental mortality) is the loss of participants.
9. *Selection-maturation interaction* can be a factor when confounding variables are mistaken for effects in the experiment.

In addition to statistical conclusion validity and internal validity, construct validity may be threatened. Carmines and Zeller (1979) define construct validity as, “[the] extent to which a particular measure relates to other measures consistent with theoretically derived hypothesis concerning the concepts (or constructs) that are being
measured” (p. 23). The main threat to construct validity is mismatching a construct and the selected study methods (Shadish et al., 2002).

According to several writers (Campbell & Stanley 1963; Borg et al. 1993; Creswell 2002; Shadish et al. 2002), factors that may jeopardize the external validity of a study are:

1. *The reactive or interaction effect of testing*, whereby the participant’s reactions to the pretest may alter the sensitivity to testing instruments.

2. *The interaction of the causal relationship with units*, whereby results may not be true if different kinds of units were studied.

3. *Reactive effects of experimental arrangements*, whereby the effects of the independent variable cannot be generalized to the person in a non-experimental setting.

4. *Multiple-treatment interference*, whereby the effects of previous experiments cannot be removed from the participant.

Often, threats to internal validity can be mitigated with the use of random sampling (Shadish et al., 2002). However, threats to validity do occur. Also, internal validity and external validity are inversely related (Shadish et al., 2002).

According to Shadish et al. (2002), compared to experimental research, quasi-experimental research may have higher external validity at a cost to internal validity. Quasi-experimental research does not employ random sampling but uses intact groups. The lack of experimental control in a quasi-experimental design increases the chances for confounding variables (Campbell & Stanley, 1963; Cook, 1979).
Face validity concerns the extent to which an instrument “appears” to measure what it is intended to measure (Borg et al., 1993).

**Educational Technology Research Methods**

Research in educational technology is rife with flawed research methods (Fleming & Raptis, 2000; Joy & Garcia, 2000; Lockee, Moore, & Burton, 2001; Lockee et al., 1999; Lookatch, 1995; Wang & Sleeman, 1993). Fleming and Raptis (2000) report that much of the research conducted in the 1980s was behaviorist in approach and lacked adequate research controls. Methodology errors, novelty effects, and the confounding of the maturation variable were among the reasons that technology was shown to be effective. Jones and Paolucci (1997) estimate that 95% of published studies on the effectiveness of technology in learning outcomes from 1993 to 1997 either had flawed designs and methodologies, or lacked appropriate quantitative measures for learning outcomes.

Bryant and Hunton (2000) report that Stickell (1963) reviewed 250 television media comparisons studies and found that 4% of 250 comparisons used sound research methodology (Bryant & Hunton, 2000). In a topographical analysis of research from 1990 to 1999, only 25% of the studies controlled for extraneous variables (Fleming & Raptis, 2000).

There are consistent and reoccuring errors in empirical educational technology research. The primary error reported in the research is low internal validity of conditions (Phipps & Merisotis, 1999; Ross & Morrison, 1996). A study is said to have low internal validity when one group has an advantage over other groups on a condition (Ross & Morrison, 1996).
Selection bias, another problem of validity, arises when the study participants are not equal between groups (Shadish et al., 2002). Many studies do not randomly assign participants to groups, which can lead to extraneous variables. Random sampling can reduce the number of extraneous variables (Bryant & Hunton, 2000; Colorado, 1988; Phipps & Merisotis, 1999).

In addition to selection bias, many studies suffer from instruments that are not reliable, or that fail to report the reliability. Carmines and Zeller (1979) define reliability as “the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials” (p. 11). The amount of random measurement error is inversely related to the reliability of the measuring instrument. Studies that fail to indicate the reliability of the measuring instruments reduce confidence in the reported results. An additional reported problem was low statistical power (Joy & Garcia, 2000; Lockee et al., 2001).

The novelty effect (that is, increased effort because of the intervention) may be another internal threat not taken into account in past studies (Bryant & Hunton, 2000; Clark & Sugrue, 1990; Colorado, 1988; Joy & Garcia, 2000). Clark and Sugrue (1990) report that increased attention leads to an increased effort which, in turn, inflates achievement gains. In secondary education studies, an intervention lasting less than four weeks had an effect size around 0.50 standard deviations; after four weeks, the effect size was reduced to 0.30 standard deviations; and after eight weeks, the effect size was reduced to 0.20 standard deviations (Clark & Sugrue, 1990; Kulik & Kulik, 1991).

The interpretation of no significant difference, part of the null hypothesis significance testing (NHST), is seen by some researchers and consumers of research as
inconclusive (Lockee et al., 2001; Lockee et al., 1999). Finally, the confounding of the instructional design and teacher effect are also threats to the validity of media comparison studies (Clark, 1983, 1994a, 1994b; Clark & Sugrue, 1990; Colorado, 1988; Lockee et al., 1999).

**Distance Training Research**

Like educational technology research, distance training research lacks quality quantitative studies (Wisher et al., 1999). Wisher et al. (1999) report the majority of studies dealing with distance training are qualitative and largely descriptive studies of users’ and instructors’ experiences. According to Wisher et al. (1999), research conducted in distance education lacks adequate experimental designs and has low internal validity. In addition, researchers in distance training generally fail to report the reliability of measurement instruments (Wisher et al., 1999).

**Costs and Benefits of Distance Training**

Web-based training has higher fixed costs than conventional classroom costs (Whalen & Wright, 2000). Web-based distance training takes less time than a traditional classroom (Hall, 1996; Whalen & Wright, 2000). Online courses are more cost-effective when the number of students is relatively high (Jung & Rha, 2000). Young (1998) reports that when mediated courses at California State University have an enrollment of 1,000 or more students, the course’s benefits exceed its costs.

Hall (1996) reports that organizations save 20 to 80% in training time when courses are delivered using multimedia technology. In training, costs are directly related to time: as time decreases, the costs of training (the costs of paying the learners to attend training) also decrease (Whalen & Wright, 2000). In 1991, the U.S. Army calculated a
savings of $1,121.00 per student using distance training technologies compared to resident training for an engineering course (Hahn, Ashworth, Phelps, Wells, Richards, & Daveline, 1991). Additionally, the U.S. Army found annual savings of $292,404 when audio teletraining replaced residence courses for unit clerks (Wisher & Priest, 1998). Schmeekle (2000) found a significant cost savings for law enforcement officers compared to traditional face-to-face instruction. The University of Ontario and the University of British Columbia report that online courses are cost-effective, with break-even points of 19 students (Jung & Rha, 2000). The Korea National Open University states that the cost of online courses per student decreased over time, dropping from $12,768 in 1998 to $7,902 in 1999 (Jung & Rha, 2000).

Little research exists regarding the cost-effectiveness of adult distributed training programs (Brent, 2001; Jung & Rha, 2000). Most comparison studies in adult distributed learning do not address cost-effectiveness or do not report the method of calculating costs.

Many different approaches to calculating the costs and benefits are found in the literature, including: the return on investment (ROI); the resource requirement model; the cost benefit model; and the cost-effectiveness model. The purpose of any costs/benefits analysis is to determine whether costs are balanced with the results of a training program (Kearsley, 1982).

Costs information in adult training usually is reported using a return on investment (ROI) approach. Human resource development (HRD) staff usually conduct ROI as part of the Donald Kirkpatrick’s evaluation framework (Phillips & Pulliam, 1999). The basic formula for calculating the ROI is to establish the benefits costs ratio
(BCR). The BCR is the total program benefits divided by the total program costs.

Program benefits usually include hard data such as decrease in time, increase in outputs, and/or increase in quality (Phillips, 1997; Phillips & Pulliam, 1999). Costs include analysis, development, evaluation, and delivery costs (Phillips, 1997).

The return on investment (ROI) model of costs analysis is very similar to the cost benefit analysis. The project costs and benefits are calculated in much the same way as the method previously mentioned. However, many ROI studies are conducted in for-profit enterprises. An ROI only includes benefits and costs that can be measured in financial terms. The process of calculating an ROI is closely related to the cost benefit analysis, in which the benefits costs ratio (BCR) is derived from the program benefits divided by the program costs. The net program benefits divided by costs multiplied by 100 provides an ROI percentage (Phillips, 1997). Since an ROI lacks evaluation of societal costs and benefits, both tangible and intangible, it is not the first choice in educational research. Moreover, the ROI method does not compare the relative cost-effectiveness of alternatives.

The resource requirement model summarizes the resources associated with a particular training approach at a given point in time. The totals can then be compared between training approaches. Total program costs include personnel costs, equipment costs, facilities costs, and materials costs, calculated over the entire instructional design of the training. The five stages of design are analysis, design, development, implementation, and evaluation (Kearsley, 1982).

A cost benefit model is the evaluation of alternative delivery methods according to costs and benefits expressed in monetary terms (Levin & McEwan, 2001). Typically,
benefits can be determined by an experimental or quasi-experimental study. If the benefits do not have a monetary value, a contingent valuation or sensitivity analysis can be conducted to determine what society is willing to pay for the benefit (Levin & McEwan, 2001). Levin and McEwan also state that observing individuals and their economic decision-making can serve as a means to provide a monetary value for a benefit (Levin & McEwan, 2001). Both costs and benefits should be discounted for multi-year projects. A BCR is established by dividing the benefits by the costs of the program. A ratio higher than one suggests that benefits outweigh costs (Levin & McEwan, 2001). A net benefits calculation can be determined by subtracting the discounted sum of the benefits from the discounted sum of the costs. Finally, an internal rate of return (IRR) can be established, which offers the advantage of not relying on a discount rate (Levin & McEwan, 2001).

The cost-effectiveness analysis model was originally employed by the aerospace industry for military and commercial systems planning. The original intent of cost-effectiveness analysis was to devise a system of statistical and mathematical rules to analyze costs in all phases of military system designs (Heuston & Ogawa, 1966). The theoretical foundation of the model is to provide a way to analyze the costs and effectiveness of alternative investments (Fox, 1965). The model allows the manipulation of variables to set an optimum level of effectiveness (Packer, 1968). The cost-effectiveness model enables researchers to compare the costs and effects of two alternative programs that possess similar learning objectives (Levin & McEwan, 2001; Patamaporn, 1992). Cost-effectiveness analysis is used when conversion of benefits to monetary values is difficult if not impossible (Kee, 1994; Laylard & Glaister, 1994;
Mitchell, 1998). A measure of effectiveness must be a valid and reliable measure, using the same principles of experimental research mentioned previously. A frequently used effectiveness estimate is the difference between the posttest and pretest scores for each treatment group and the control group (Levin & McEwan, 2001). For each alternative, the costs and effectiveness estimates are used in a cost-effectiveness ratio (CER), where cost (C) of a given alternative is divided by effectiveness (E) (Levin & McEwan, 2001).

The cost-effectiveness ratio is the cost requirement to obtain a single unit of effectiveness. When two alternative programs are compared by CERs, the lowest ratio is the more cost-effective. If estimates of costs or benefits are likely to involve uncertainty, a sensitivity analysis can be used to display CERs with different assumptions (Boardman, Greenberg, Vining, & Weimer, 1996; Levin & McEwan, 2001). The sensitivity analysis should include a high, middle, and low estimate of the cost or effectiveness estimate in question. A cost-effectiveness analysis enables researchers to study secondary outcomes as well.

**Defining Costs**

According to Levin and McEwan (2001), the costs of an educational intervention are the total value of all resources given up, had they been assigned to their most valuable alternative use. There are three primary steps in determining costs for any cost model (Jamison, Klees, & Wells, 1978). The first step is to collect historical data (components) of all costs involved and decide what costs are to be included (Jamison et al., 1978; Kearsley, 1982, 1997; Phillips, 1997). Resources or components that contain costs include: personnel costs (listed by their roles); facilities (the physical space required for
the program); equipment and materials; and other inputs not categorized in the previous three components (Levin & McEwan, 2001).

For programs lasting more than a year, costs should be discounted and adjusted for inflation (Brent, 1996; Kearsley, 1982; Levin, 1985; Levin & McEwan, 2001; Rouse & Boff, 1999). Costs are discounted because investments are worth more than the same amounts gained later (Rouse & Boff, 1999). Costs in the future are less burden than costs at present (Levin & McEwan, 2001). A dollar earned today is worth more than a dollar earned a year from now because of the resulting interest the dollar would earn (Phillips, 1997). According to Brent (1999), costs are compared in present units and not future units “because we may not be around in the future to benefit and because interest can be earned on the current time called compounding.”

For example, if the interest rate is 10 percent, the current unit will be preferred to the future unit because it will be worth \((1+i)\) or 1.10. To compare a present unit to a future unit, it is necessary to discount the future unit. Next year’s amount must be divided by \((1+i)\) to obtain the present value.

An interest rate or social discount rate is used to determine the interest that would have been paid on the capital (Rouse & Boff, 1999). According to Levine (2001), there is little agreement on the discount rate to be used in cost effectiveness analysis for social research. Therefore, a sensitivity analysis using rates from 0% to 10% may be used. Finally, costs should also be adjusted for inflation. Costs that occur over a multi-year project should be adjusted according to a predetermined consumer price index (CPI) for a certain year (Levin & McEwan, 2001). Each year, costs should be adjusted using the same year CPI.