An Examination and Critique of the Theoretical and Empirical Literature on the Use of Diagrams, Graphs and Other Visual Aids in the Learning of Scientific-Technical Content from Expository Texts and Instruction.

James Carifio and Rocco J. Perla
University of Massachusetts Lowell

There is no first author or second author for this article. Each author is first author and has contributed equally to this article which is an inter-disciplinary work.

Abstract

This paper critically reviews and analyzes key studies that have been done in science education and other areas on the effects and effectiveness of using diagrams, graphs, photographs, illustrations, and concept maps as “adjunct visual aids” in the learning of scientific-technical content, and those studies that have students draw diagrams, graphs, maps, and charts to express their understandings of the concepts and relationships that are present in the text they read or/and empirical data provided (i.e., student-generated “adjunct visual productions”). Visuals and visual aids of all kinds are a very important part of science and science education and increasingly so with the growth in scientific complexity in the last fifty years. In general, the research and “theory” on instructional aids is fragmented and somewhat unsystematic with several flaws and a number of key uncontrolled variables, which actually suppress and mask effects in the studies that have been done. The studies that have been done for the most part have been “functional,” “pragmatic,” and practitioner-oriented studies, which may reflect to some degree the difficulty and complexities of the research and theory construction problems in this area. In this paper, the findings of these studies are compared to relevant literature and empirical research and findings in the areas of cognitive psychology, computer science, neuroscience and artificial intelligence that help to clarify many of the inconsistencies, contradictions, and lack of effects found for visual (e.g., diagrams and graphs) instructional aids in the science education literature currently and in the past twenty years. A model and a set of criteria and goals for improving research in this area is described, as “visuals” are a first step in the process of learning formal (scientific) models, which are most often visually represented. Little is known about how students learn formal models, which is the real research challenge of the next twenty years, both within and outside of science education.

Introduction

There is an abundance of research in the science education, educational psychology and instructional theory and technology literature on “aids” to enhance learning from expository text. These “aids” are “visuals” of all kinds, advanced organizers, concept maps, pre and post questions and tests and other adjunct instructional activities ranging from “drill and kill” exercises to simple “home lab” experiments and mini-projects. Since the late 1960’s, there have been a wide variety of models and “theories” posed about instructional aids as well as each of the aforementioned type of aid specifically (e.g., Frase, 1967; Ausubel, 1976; Paivo, 1986; Kosslyn, 1994; DeVega, Intons-Peterson, Johnson-Laird, Denis & Marschark, 1996; Byrnes, 2001). In general, however, both the research and the theory on instructional aids is fragmented and somewhat unsystematic and in most instances “functional,” “pragmatic,” and practitioner oriented, which may reflect to some degree the difficulty and complexities of the research and theory construction problems in this area.

Lack of standardized terms and concepts, weak theoretical foundations and, in some cases, questionable research designs have made the literature in this area inconsistent, difficult to evaluate, and particularly difficult to extend and/or operationalize (see Levin, Anglin, & Carney, 1987; Ruiz-Primo & Shavelson, 1996). This difficulty has been particularly true in the area of science education and the learning of scientific technical material, where the quality of the empirical research done, it may be argued, has declined over the past two decades with the decline of strict experimental and scientific research (see National Research Council [NRC], 2002). Many of the most basic elements of “instruction,” such as what is learning, what type of learning and learning levels are being investigated, what is a text (theoretically and in fact) and how is it processed, what is a visual (theoretically and in fact) and how is the “visual” represented, processed, and then assimilated and integrated into long term memory schemas are undefined, not addressed, or at best fuzzily addressed in this research literature.

Part of the problem and irony is that instructional aids research, and particularly research on “visuals aids” in science education is that it is and has not been “good science” relative to modern definitions of the term (see NRC, 2002) or “good experimental research” (see Kerlinger & Lee, 2000). Further, a great deal of sophisticated and empirically supported progress that has been made in the last twenty years relative to the aforementioned instructional elements and questions has been made in the areas of academic cognitive psychology, computer science, neuroscience, and artificial intelligence (see below). These areas and foundational disciplines, which are not often included and operationalized in the science education literature and research (see Fensham, 2004; Perla & Carifio, 2006) are now the real “mainstream” of the theoretical and empirical work on these instructional elements and questions (Ashcraft, 2002). The important point to be made here, then, is that the science education literature on the “visual aids” question is extremely noisy and not very helpful in making instructional decisions, or in interpreting the results of studies that have been done in science education concerning their efficacy or effects.

This paper, therefore, examines and critiques some of the relevant research that has been done in science education and related areas on these “adjunct instructional aids” questions and the effects found, or not found, when they are used. It then compares these studies and their findings to the relevant literature and empirical research and findings in the areas of cognitive
psychology, computer science, neuroscience and artificial intelligence that can help clarify many of the inconsistencies, contradictions, and lack of effects found for visual (e.g., diagrams and graphs) instructional aids in the science education literature currently and in the past twenty years. Further, the findings and insight from the latter domains and fields of study, it will be shown, may also be used to guide text and instructional materials creation decisions during both development and field testing (see Perla & Carifio, 2005 for details). The second article of this series then outlines a conceptual model, theory and a set of criteria and goals for improving research in this area of instruction, as “visuals” are a first step in the process of learning formal (scientific) models, which are most often visually represented.

**Representational Systems**

One of the central problems with the majority of the literature and studies reviewed below concerns the overly simplistic behaviorist view and assumption present in most of them that text and language are strings of words (or homogenous units representing homogenous structures) that are somehow fundamentally different from “visual” representational systems such as diagrams and graphs. This view is diametrically opposed to the mainstream research in cognitive and computer science (e.g., Neisser, 1967, 1987; Norman, 1981, 1982; Selfridge, 1988) and the area of artificial intelligence (e.g., Minsky, 1975, 1986; Pitt & McCulloch, 1988) which recognizes that verbal, mathematical and visual systems (and displays) are all forms of representing (deep structure) information, and all have their own but similar language. Further, modern cognitive psychology and the cognitive sciences contend (and have vividly demonstrated) that all visual information and stimuli may be represented propositionally (as with computers) and at different levels of depth (e.g., Norman, 1981; Minsky, 1986; Zhang & Norman, 1995). It is the propositions that define the relationships in a deep knowledge structure and/or surface display (or topology) and provide the semantic (meaning and interpretation) component in any system (Penrose, 1991). This propositional representational view is in fact the central and core postulate and contention of modern cognitive learning theory (Ashcraft, 2002; Norman, 1981; Selfridge & Neisser, 1995). In many of the studies described below, only the surface structure of the text is considered and examined (and only weakly so), while the quality of the conceptual, causal and explanatory (deep) structure of the text is not addressed at all. Contemporary (mainstream) models of text processing and reading comprehension, it should be noted, have shifted from analyzing surface structure to analyzing deep (macro and micro) structures and their role and function in learning and understanding (e.g., Dagostino & Carifio, 1994; Kintsch, 1986; Graesser, Swamer, Baggett & Sell, 1996; Meyer, 1986).

Gestalt psychology and cognitive theory have convincingly demonstrated that “visuals” depict, show and establish relationships and that all learning is the learning of relationships (Neisser, 1967; Ashcraft, 2002). However, “visuals” do not have and do not communicate the key and fine-grained details, specifics and nuances of relationships. For example, there is no set of simultaneous equations that can represent any knowledge base beyond a simple Euclidean proof and none that can represent a knowledge base that is dynamic and generative. Any equation or mathematical algorithm or model must include semantic text to mean anything (Zhang & Norman, 1995). Pictures, diagrams and graphics are also more ambiguous and primitive representations (i.e., the iconic level to Bruner) and cannot really represent cognitive networks and semantic webs effectively or efficiently. These points are why and how “visuals”
are impoverished and limiting, and why contextualizing and rich text (or/and dialogue) is needed, and why both are always present in something deeply learned in an integrated fashion, as opposed to chunks and categories of fuzzily related information or images, or both simultaneously if one holds Paivo’s (1986) dual (verbal and spatial) encoding model, or DeVega et al.’s (1996) multiple encoding model. All of these latter points are also why items such as concept maps (see Ruiz-Primo & Shavelson, 1996; Mintzes, Wandersee, & Novack, 1997), and even very elaborate and detailed concept maps, represent only a very small and impoverished amount of the information in the actual knowledge structures that they “depict,” and that there are several missing levels of (much more detailed) representation before one gets to the deep knowledge structure in question (see Norman, 1983; Carifio, 2005). In a word, illustrations, diagrams, charts, tables, graphs, concept maps and other such visuals displays always have a high degree of “excess” and “ambiguity” as well as “inherent” and “latent” meaning (even to experts), which cannot be reduced to the level of “noise” except through accompanying text and a knowledge of the conventions, language and semantics of the visual display in question. Visuals, therefore, may convey a very great deal of information (and particularly relational information) almost or close to “instantaneously” and “more transparently” (particularly for relationships), but at the very high price of “excess, ambiguous, and latent” meanings which need operationally narrowing, defining, and often highly nuanced (thick and rich) text to reduce the “noise level.” Textual and visual representations are, then, somewhat like quarks; namely, they naturally come in pairs and to separate (de-contextualize) them significantly is to more than abuse them. It is to both seed and fuel a whole host of potential unintended and unwanted consequences through unwarranted and often unconscious inferences, deductions and interpretations, which are broadly referred to as misunderstandings, misconceptions or alternative conceptions.

It is often said that one picture (or visual) is worth a thousand words (of text); but visuals without text are most often just “pretty pictures,” and pretty ambiguous pictures as well. Further, the point that is rarely made (and particularly as a counter argument) is that one word (e.g., freedom, democracy, friend) is often worth a million plus pictures (as any poet or novelists knows), and that text is far more powerful and generative than visuals because of the richness and layers of nuanced semantic networks, a point that is often forgotten in the current “moving image” and “sound-bite (i.e., impoverished text)” age we live in or in what is now referred to as the “presentation culture” (mostly image and little substance) by current sociologists (see Higgins, 2007). And the power of text is even fading from unconscious memory as trade books, magazines, textbooks, educated TV programs (and MacDonald’s cash registers) are purposefully increasing the amount of space/time occupied by visuals as compared to text, with the stated rationalization that they are “crafting communications” for the image and visually conscious millennial generations as this generation is not as (textually) literate as previous generations, as Hirsch (2006) has documented with a vengeance and recent research on the reading habits and behaviors of Americans by the National Endowment of the Arts (2007) has shown. Visuals and visual aids in learning, therefore, are even more important in instruction today than they are made out to be by science educators and those doing research on the effects of visuals and visual aids in the learning of scientific-technical content. To elucidate the above points and others, then, we begin by considering illustrations and diagrams in science texts.
Illustrations and Diagrams

In exploring the use of *illustrations in science text*, Mayer (1989) studied 34 female college students who lacked prior knowledge about auto mechanics (self-reported) using a posttest only design. However, the characteristics of an illustration are not defined and the criteria for a good or bad illustration are not described in this study, which is typical of these studies in the science education area as well as other areas. In other words, the validity and quality of the independent variable is not addressed in this study (and similar studies), but is only inferred through the results of the study, which creates a circular, unreliable and psychometrically dubious method and experimental design. The specific mechanical concepts in this study involved the hydraulics of drum bakes. Students were randomly assigned to the “text with systematic illustrations” group or the “text with no illustrations” group. Mayer predicted that illustrations would improve explanatory recall (recall of explanatory text that systematically described system function), but not non-explanatory recall (no systematic descriptions of system function), because the “systematic illustrations will help direct the reader’s attention towards explanatory information and away from other information in the text” (p. 243). As predicted, the illustration group demonstrated greater gains in explanatory recall (28%) versus the no illustration group (16%). However, a supplemental t-test found this difference to be marginal (p < 0.6). This marginal difference could have resulted from illustrations of poor quality, which would neutralize, reduce or eliminate the purported “effect” of the illustration, or simply the illustration group having more information overall than the other group as well as having spent more time-on-task studying the illustration.

Mayer also predicted that the systematic illustration group would help students construct coherent mental models and that these models would help students answer transfer (problem solving) questions that require “running” and “executing” one’s own model to answer questions posed. It was also assumed by Mayer that these models should not result in increased verbatim recognition of the factual material in the text since the text material could be memorized directly from the text or organized around the model. Mayer argues that the results support this contention as the responses from the illustrations group to the transfer posttest questions were significantly more creative and effective than the no illustration group (p < .01), but no significant difference was found between groups for verbatim recognition. Although statistically significant, the data is difficult to interpret meaningfully since the creativity (“transfer”) scoring measures and procedures, although explained in detail, are not well linked to any theoretical constructs of creativity or related traits claimed. Further, no reliability or validity data are reported for the scoring instrument.

In a follow-up experiment in the same study, Mayer sought to understand if the differences observed in explanatory recall and problem solving transfer were attributable to the labels on the diagrams, which restate the key explanatory information, the graphics, or the combination of the two. As with the term “illustration” in the previous study, terms like “diagrams” and “graphics” are not clearly defined. The experiment measured explanatory recall and problem solving transfer by comparing three groups: illustrations, illustrations without labels, and labels without illustrations. Compared to the other groups, the illustrations group recalled substantially more explanatory information, however an ANOVA found no significant
differences between groups for non-explanative recall, although it is unclear what non-explanative recall was exactly. The illustration group also performed better than the other two groups \((p < .05)\) relative to creative problem solving. An ANOVA followed by a supplemental Newman-Keuls test found that the illustration group performed significantly better than either of the other groups \((p < .05)\). However, it should be noted that this post-hoc test is much more liberal than other post-hoc tests relative to finding post-hoc differences (Kerlinger & Lee, 2000). No significant difference between groups was found for verbatim recognition. The results suggest that labeled diagrams (including both graphics and labels) help students construct meaningful mental models, since they appear to facilitate explanatory (versus non-explanatory) recall and creative problem solving (versus verbatim recognition), respectively. Why there are or would be no differences between the treatments in verbatim recall is left unexplained in these studies, which is a problem both theoretically and empirically. No effect sizes were reported for these studies nor detailed descriptions of the characteristics of the subjects in these studies, which is an important factor as will be seen below.

This research did not consider the fact that students in the text with illustrations group received more external stimuli, instruction, and information than students in the text alone group. Cognitive learning theory predicts that additional learning stimuli, instruction, and information (regardless of its specific form) should aid student understanding (Ashcraft, 2002). This research also employed, in our opinion, a behaviorist view of text (and language), and it sees both (text and language) as being strings of words (or homogenous units representing homogenous structures) that are somehow fundamentally different in kind and character from “visual elements.” Mayer actually sees the (textual/verbal) labels he put on the illustrations as Skinner-like “signals,” “mands,” or “cues” (Skinner, 1957) for directing attention, orienting the viewer and reducing “completing (interpretative) responses,” rather than as elements that fuse textual and spatial representations together (making “quarks”), which trigger arousal of semantic long term memory and knowledge leading to deeper processing of the visual information and full dual encoding and processing of the fused information. But even more importantly, Mayer treats the text in these studies as if it has no structure and completely ignores any effects the structure of the text may be having on the learner and learning process (see van Dijk, 1980, and Kintsch, 1986 for details) because he is extremely over focused on the illustration in the sea of competing instructional components and explanations. In this respect, Mayer is not atypical of the science educators doing research on this topic and these questions as well as researchers in other areas including psychology and computer science.

It is also important to point out that Mayer’s studies assumed that less skilled learners in a particular domain have the most to gain from meaningful instruction (i.e., instruction that encourages the construction, conceptualization and use of mental models) because these types of learners are less likely to spontaneously construct such models compared to high skilled learners (see Snow & Lohman, cited in Mayer, 1989). That being said, a major drawback and limitation of this group of studies is the fact that Mayer did not report the ability and developmental levels of his subjects, nor did he analyze his data accordingly. It is likely that developmental and reasoning levels, which vary significantly among college students, influenced student performance on the assessment activities, aside from their background knowledge and experience with the content in question. Further, the structure of the (encyclopedia-derived) text is characterized only cursorily in these experiments as a “sequence text structure” that depicts “a
series of state changes...in chronological order” (Mayer, 1989, p. 241). This characterization is almost a verbatim quote of Skinner’s definition of an instructional sequence (Skinner, 1958), and to Skinner (even after his “great debates” with Chomsky), there was no such thing as deep or latent structure; there were only surface topologies (“what you see is all there is, never mind what you get”). Whether students also interpreted the text in this linear and sequential fashion is difficult to ascertain from this study and can only be weakly inferred through the data, but work from a wide variety of studies has shown that this tends not to be the case, and particularly so with better students who tend to read and process information in a wide variety non-linear patterns (see Carifio, 1978), which have also been observed in the study of sciatic eye movements in the processing of visual information (see Ashcraft, 2002). Neither the superficial or deep structure and quality of the text is addressed, explored or validated in this study, which leads one to believe that Mayer does not view text as a complex, dynamic and sophisticated representational system or explanatory structure that can be well or poorly developed. In addition, the actual scientific diagrams and illustrations (although derived from text books) were not validated for the purposes of these studies. Again, we have found that these weaknesses are typical and not atypical of research on this topic and these questions in the science education area and in other areas as well (Perla, 2006).

In a follow up study to the above research, Mayer and Gallini (1990) had four groups of 24 college students read expository descriptions related to three scientific systems (brakes, pumps and generators). The four treatment groups included students that (1) read a booklet on braking systems with no illustrations, (2) read a booklet containing illustrations representing major parts of each type of braking system, (3) read a booklet demonstrating the major actions and steps of each braking system or (4) read a booklet containing both parts and steps illustrations. No inter-rater reliability or internal validity data were presented for any of the four treatments in this study. In each group, half of the students were deemed to be low prior knowledge students and half were deemed to be high prior knowledge students via self-reporting. Aside from this self-reporting of knowledge-levels by students, no other criteria were used to distinguish high and low prior knowledge students. The classification “high prior knowledge” was also somewhat misleading as these were students “who rated their knowledge of automobile mechanics as more than ‘very little’ and reported having performed minor car maintenance” (Mayer & Gallini, 1990, p. 718). So the level of prior knowledge—which was a focus of this study—was only weakly addressed, not directly measured, and not psychometrically characterized or validated.

As with the previous two studies, developmental level was not accounted for in this study. Analysis of variance and supplemental testing procedures “indicated that the parts and steps (but not other) illustrations consistently improved performance on recall of conceptual (but not non-conceptual) information and creative problem solving (but not verbatim retention), and those results were obtained mainly for the low prior knowledge (rather than high prior knowledge) students” (Mayer & Gallini, 1990, p. 715). The authors note that the major difference in this study between the effective explanatory illustrations (parts-and-steps illustrations) and the other ineffective illustrations, was that the former visually portrayed both system typology (step illustrations) and component behavior (parts illustrations), while the latter portrayed only one of these two elements. It should be noted that the parts and steps strategy portrays, elaborates and describes more relationships and more kinds of relationships than the
other strategies employed in this study, which is consistent with and predicted by the view and characterizations of “visuals” we gave at the beginning of this article. As stated previously, a central feature of cognitive learning theory, and the construction of meaning, is emphasizing the number, type and completeness of relationships and representations. Although these findings are reasonable and intuitive, the fact that none of the independent variables (or treatments) was subjected to external, independent panel validation of their alleged characteristics or lack thereof requires that these data be interpreted very cautiously (see Lanza & Carifio, 1991 for details).

In extending the work of Mayer and colleagues, Hall, Bailey and Tillman (1997) examined whether verbal information could be used by college students to generate a visual representation. It should be recalled that the illustrations and diagrams used by Mayer and colleagues were not student generated. In the Hall et al. study, 92 freshman college students read the same text passage that described how a hand air pump worked. Three different instructions to the text passage were developed and each served as an experimental condition. The three different instructions included a text condition (students read text with no illustrations), a picture condition (students read text with illustrations), and a generate condition (students read text and draw a representation of the pump). This model basically compares passive (first two conditions) versus active (last condition) learning and responding. These student drawings may be termed student-generated visual productions, which may be seen as “aids” as in “facilitators of learning” or seen as dependent variables as in “outcomes of learning.” Both are different from but variants of “researcher or professional constructed presentation visual aids.” These differences and distinctions as well as similarities should be kept in mind as both kinds of visuals are used and often used in the same study (and in everyday instruction), such as the study being discussed now by Hall et al. (1997). With regard to students ability to generate accurate representations from text, significant main effects for the subtest scores (p < .01) and total scores (p < .05) were found for the three conditions. Tukey post-hoc tests revealed significantly higher scores for students in the generate condition compared to students in the text condition. No other significant differences between means were found. As with the studies mentioned above, individual student differences were not addressed.

For example, student developmental levels and ability to visualize are excluded from the analysis despite the fact that reasoning ability and the ability to visualize differ across the population and have been shown to be a source of significant difference in numerous other studies summarized by Kosslyn (1994) as well as in the ability to do mental imaging (e.g., Kuzendorf, 1981; Ashen, 1992; Kuzendorf & Wallace, 2000). Hall et al also point out that had their study been designed to replicate Mayer and Anderson’s (1991) study to include (by collapsing their design) just the same subtests and two conditions (“text” and “text and picture, provided or generated”), significant main effects were found (p < .05). However, as with the studies by Mayer and colleagues, the structure, features, and quality of the text or the picture were not addressed, and the pooled “effect” is likely due primarily to the active and constructivist responding of the students who generated the “picture.” In terms of the “picture” or the “visual” used in the study, something as simple as how many elements and relationships it contained would be extremely useful information for gauging and comparing studies and locating visuals along complexity continua, and in terms of their degrees of embeddedness; namely, additional abstract, objective and theoretical properties of the “visual” being studied. As
In contrast to the work of Mayer and colleagues that addressed experimenter provided illustrations, and the work of Hall et al (1997) that addressed student-generated illustrations, Van Meter (2001) examined the technical accuracy and cognitive influence of student drawings relative to a unit on the central nervous system that received varying levels of visual aids. Van Meter’s subjects were fourth and fifth-grade students randomly assigned (within grade level) to one of four conditions including a (1) reading condition (read) where learners read the passage and inspected a provided illustration, (2) drawing condition (draw) where learners were told to read each page of the text and draw a diagram to represent what was learned for each page, (3) illustration comparison condition (IC) where learners executed the same task as in the drawing condition, but were told to inspect and compare their drawing to a provided illustration and to make any changes they thought were appropriate and (4) prompted illustration comparison condition (PIC) was the same as the IC condition, but the learners were asked prompted questions to help analyze the quality of their drawings. These “prompted questions” were intended to focus students attention on shortcomings in their illustrations and students were allowed to make changes to their illustrations based on these questions. The PIC condition is consistent with the “guided-discovery” approach and method of the 1960’s and 1970’s.

Using multivariate and one-way ANOVA followed by Tukey’s HSD post-hoc procedure on a host of dependent measures, results demonstrated that (1) the PIC cohort (guided discovery group) constructed significantly more accurate drawings than either the draw or IC cohorts, (2) the PIC cohort scored significantly higher on the free-recall measure than the read cohort, (3) the PIC cohort participated in significantly more self-monitoring events compared to the draw cohort, while the read cohort participated in fewer self-monitoring events than any other cohort, and (4) the PIC cohort had significantly more time on task than the draw cohort, while the read cohort spent less time on task than any other cohort. In general, these findings are consistent with the established view that the use of external aids and instructional supports (or guided discovery) can enhance student performance, accuracy and understanding (Gagne, 1965), that performance with feedback is superior to performance without feedback (Page, 1958; Thorndike, 1935), and that “highly active (normatively)” and (meta-cognitively) “reflective” behaviors produces ranges of superior outcomes whether these processes are self-regulated or instructionally (environmentally) regulated (e.g., Polya, 1965; Pajares, 1996). The Van Meter (2001) study, therefore, is about much more than the effects of visual aids, but this point is actually a “good thing,” as these other instructional components serve as control (comparison) and falsification points, and give estimates of the net effects of multiple components, which is more typical of actual instruction and instructional practice.

It should be pointed out that the above study was of fourth and fifth graders, but that the treatments administered were more consonant with the “learning profiles” of older students and adults given the core views of both cognitive and developmental learning theory. One would, therefore, expect (according to these theories) a more pronounced effect for the “guided discovery” visual and text processing treatment with adults, and with adults who had not developed the self-directed and meta-cognitive processes and skills that the guided discovery (processing) treatment provided, again with the caveat of not having any specific and concrete
descriptions of the complexity of the visuals or the text. The above study, however, illustrates the effects these variables (text, visual aids and guided processing all together) have on young students, which demonstrates the points we made previously about instructional elements being “quarks (fused elements),” but particularly so with text and visual elements. The above findings also suggests that aspects of utilizing visuals and text, and both together, are in part learned behaviors and that some students may be well schooled in these representations and types of processing, and others may not be by the time they are in high school and college, and this basic fact may be another factor in the problem and assessment of effects in this area. Differences in meta-cognitive skills and self-regulated learning processes are additional uncontrolled variables in studies of the “visual aids” question, along with the other variables identified previously. The importance of meta-cognitive and self-regulated learning skills is illustrated in the next study reviewed.

Using instructional examples from the life sciences, Schwartz (1993) conducted two experiments that examined whether adolescents were capable of constructing abstract symbolic visualizations in order to structure complex information. In this study, symbolic visualizations were defined as “visualizations that do not resemble their referents … [namely,] abstract, non-sentential (without sentences) representations of structure” (Schwartz, 1993, p. 1309).

In the first study, 7th and 9th/10th graders randomly received one of eight experimental (instructional) packets. Four of the packets covered topics that were “familiar” to students (e.g., food webs) and four packets covered “unfamiliar” topics (e.g., protein pathways). Each packet contained a page describing the basic concepts to be studied, a second page to construct visualizations (diagrams) and a third with a reading comprehension question and two inferential questions based on relational sentences on page one. Some students completed the activities individually while others worked in dyads. Briefly, the visualization scoring rubric involved three general categories including (a) relationship direction indication (i.e., none, implied or explicit), (b) one-to-many relations (scored yes/no) was used as an indicator of whether a visualization included structural aspects from each of the three categories, and (c) many-to-one relationships, again scores yes/no and used as an indicator similar to category (b). Using this scoring scheme, two independent coders achieved 97% agreement.

Results showed that 66.2% of all visualizations included structural elements in each of the three scoring categories (chance estimate = 22%). A MANOVA found no significant effects with regard to topic familiarity, grouping, and no interactions. Grade level (which was a surrogate for level of cognitive development) was the only factor producing a significant effect (p < .01), which is one of the findings that lead us to point out the importance of cognitive development level as an uncontrolled intervening variable in the studies of visual aids in science education (besides developmental theory itself). This “grade level” effect was located in the many-to-one category using univariate analysis (p < .01). In this case, the upper grade levels (9th/10th graders) most often used path diagrams to represent the many-to-one structure, while 7th graders tended to use letter repetition strategies to represent this structure. These results suggest that students can structure novel and complex information through visualizations (a very important finding) when directed to do so (the caveat) and that developmental level may affect student performance (the key qualifier of the finding). Again, no information of the complexity of the visual and visualized elements were reported relative to the number of elements and
number of relationships, so it is difficult to assess whether the visual and visualized elements were simple (with baseline effects being observed) or very complex (with threshold or “too difficult” effects being observed). The findings from this study, however, raise questions about the validity of results obtained for other studies looking to assess student’s conceptual understanding using “visuals” and visual materials that did not account for developmental level, as well as a number of other uncontrolled variables. Further, it should be pointed out that a portion of the general population cannot visualize very well at all, and that a portion of the population is superior at visualization and imagery (e.g., Ashen, 1992; Kosslyn, 1994; Kuzendorf & Wallace, 2000). These types of established individual differences must be addressed, because they produce effects or lack of effects, which correlate with other differences that are buried in the design of the present studies and future studies. Therefore, in reality and fact, a picture (i.e., “visual”), when it is worth a thousand words, is only worth a thousand words for some students (i.e., roughly 80%) and not all students (i.e., roughly 20%) for whom text is actually better and worth more than the visual. This fact is an important fact and interaction effect that all of these researchers have missed or ignored. This fact and interaction effect is not only true about “presented visuals” that accompany text, but is also equally true about student generated “visual productions” of text concepts and relationships, which has been even more “ignored” by those who use this methodology to measure such dependent variables, never mind the great difficulties that exist in the scoring and indexing of these productions (see Ruiz-Primo & Shavelson, 1996 for details). These differential effects (interactions) should be kept clearly in mind when reading and interpreting all of these studies and in designing and conceptualizing future studies.

Based on findings from Experiment 1, Schwartz (1993) hypothesized that the more frequent use of path diagrams by the 9th/10th grade students was the result of analogical transfer from previous instruction. Using a two treatment, pre-post experimental design, Experiment 2 by Schwartz explored whether 7th grade students with average ability could (without cues, or more accurately, semantic long term memory arousal) spontaneously transfer a visual problem solving strategy (taught during the instruction period) to a novel problem. Table 1 is a recreation of the experimental design used in Experiment 2. As can be seen in Table 1, the only difference between experimental groups is whether Path Diagrams or Cartesian Diagrams were taught during day 16 of 17 days of instruction. This design could be said to be looking for and claiming effects from an overly narrow and very small amount of the total instruction given, but qualitatively the difference in enabling capabilities between these types of diagramming techniques may be powerful enough as to produce significant effects, and do so independently of the other 94% of the treatment. The pre and postest scoring was based on whether the student made a path diagram, a different type of visualization, or made no visualization, with a two coder agreement of 100% necessary on the classification for it to be accepted into the analysis sample.

Pooled results across treatments suggested that students adopted a visualizing strategy after instruction (pretest visualization, 2.6%; posttest visualization, 52.6%) and that this was a reliable difference at the .01 level. However, this outcome could be a sensitization effect, as well as a regression effect (particularly given the very large gain observed), and these two factors together are a competing explanation that was uncontrolled in the design of the study. Despite
Table 1. Experimental design for the Schwartz (1993) study, Experiment 2 (From Schwartz, 1993, p. 1317).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Test</th>
<th>Instruction</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>PATH</td>
<td>Pretest</td>
<td>Matrices</td>
<td>Path Diagram</td>
</tr>
<tr>
<td>No-PATH</td>
<td>Pretest</td>
<td>Matrices</td>
<td>Cartesian Graphs</td>
</tr>
</tbody>
</table>

The fact that more students in the Path group engaged in visualization during the posttest (64.3%) compared to the No-Path group (45.8%), this difference was not significant (p > 0.25). The author contends that the results support the hypothesis that prior instruction in and knowledge of path diagrams was analogically transferred to the problem-solving task, as there was a 50% posttest gain in the frequency of path diagram usage in the path treatment group and only a 8% increase in path diagram use in the No-Path group, a difference that was significant at the .01 level. However, these types of results and differences are consistent with meta-analyses of pre-test sensitization effects, although the post-test in this study was at the end of the three week period when sensitization effects tend to “wash out,” which would suggest that the unit itself was the actual “test” (see Kerlinger & Lee, 2000) and is the effect one is seeing in the “transfer test.” The results also showed that students who used visualizations (path and no-path groups) demonstrated problem-solving gains in the posttest that exceeded gains made by students who did not use visualization to solve the problem, but this difference was not found to be statistically significant, but most probably would have been if individual differences in visualization and developmental levels had be measured or estimated and used in the analyses. Although the author notes that this study was not designed to address problem-solving effectiveness, even raising this point is problematic since it conflates and confounds problem solving with parallel forms of instruction, which is a more than tenuous comparison.

Following a similar line of inquiry, Gobert and Clement (1999) studied 58 fifth-grade students by examining the effects of student-generated diagrams versus student-generated (written) summaries on conceptual understanding of causal and dynamic knowledge of plate tectonics. While reading expository text, students were asked to (a) draw diagrams, (b) produce written summaries, or (c) read the text (control). The researchers hypothesized that student-generated diagrams would facilitate conceptual understanding by promoting inference making (Schwartz, 1993), by providing perceptual cues and spatial relations (Larkin & Simon, 1987) and by explicitly indicating structural relations between variables and concepts (Brueker, 1984). The text based instructional model used in this study followed the work of Kintsch (1986) and Johnson-Laird (1983) that assumes that “higher-level inference tasks are best supported by representations which reflect higher-level, more integrated representations—that is, situation [“situated”] models…or mental models” (Gobert & Clement, 1999, p. 41).

The assessments in this study involved having students in each group mentioned above perform intermittent (during reading) tasks, which included four separate domain-specific measures, and then answering multiple choice and short answer questions as well as executing a diagram task in a posttest booklet. Coding schemes were developed to assess intermittent and post test data. Inter-rater reliability with a second coder was reported for the posttest diagram coding scheme (93.8%). No reliability coefficients were reported for the short answer or
summary responses. With respect to tasks that were executed during the reading of the text, students who generated summaries outperformed students who generated diagrams on each of the four intermittent measures, and the results of two of these measures were statistically significant at the univariate and multivariate levels of analysis. The intermittent data was also pooled into spatial/static and causal/dynamic measures of the content domain with the summary group again outperforming the diagram group on both measures, although statistical significance was only found for the spatial/static measure (p < .001). This trend, however, was reversed during the posttest analysis, which was intended to measure students’ conceptual representations or mental models of the domain. During the posttest, a univariate test demonstrated that the diagram group had a significantly better understanding of static/spatial knowledge (p = .017) and causal dynamic knowledge (p = .018). This finding, it should clearly be noted, can be explained by the fact that writing tasks tend for most people to more strongly arouse and engage episodic long term memory than semantic long term memory, whereas diagramming and visualizing tasks tend for most people to arouse and engage semantic long term memory than episodic long term memory (Ashcraft, 2002). This differentiated view of human long term memory and information processing is not the view of these two phenomena (semantic and episodic long term memory) that guided the Gobert & Clement (1999) study, or were its implicit theory as will be seen below. It is somewhat obvious, given the results of the Van Meter (2001) study discussed earlier that one wants instruction that actives and engages both types of memory in actual instructional situations and one should remember that Gobert and Clement (1999) were trying to establish a particular experimental (and opposed to instructional) point and finding.

Based on the above findings, Gobert and Clement (1999) contend that the student-generated summaries did not encourage high-level mental model representation, and that the summaries were more of a rote recall task (which will come as a surprise to most language arts teachers), even though being able to summarize material is classified as (higher order) comprehension in Bloom’s taxonomy. Conversely, student generated diagrams encouraged the formation of elaborate mental models, but did not aid in the development of a similar conceptually oriented text-based representation (memory of the text itself). This finding could be due to poor conceptual presentation structure (order) of the text itself, as studies by Frase (1969), Shavelson (1971) and others (e.g., Anderson, 1967; Norman, 1981; Kintsch, 1986) have shown that students recall (memory) of a text tends to strongly resemble the conceptual presentation order of the text to be remember and recalled. However, it is important to emphasize that these findings clearly suggest that both summaries and diagrams are important factors in learning, as each seemed to encourage different and needed forms of understanding, and are “quarks,” as stated previously, that work together to make better and more elaborate and detailed memory schemas and knowledge structures. Also, without knowing the nature, structure and level of the text (treatment) used in this study, it is difficult to know the “how or why (process or explanations thereof)” of the types of summaries or diagrams that students “translated” from the text. And once again, no information was provided on the complexity of either the diagrams or summaries created (number of elements and number of relationships) that would have helped to contextualize the results of this study and compare them to the results of others studies. To conclude, as these authors did, that diagrams alone were more important than text-based summaries for the elaboration of mental (conceptual) models is more than somewhat odd considering their results and that the function of text is elaboration and explication (and
particularly of diagrams), and the fact that diagrams are a more primitive form of representation than text.

It is important to remember that not all text is the same or equivalent, which is more a behaviorist rather than cognitivist view of text. It is also important to remember that text in a very large number of these studies is not “thick and rich” real word text, but highly artificial and simplified text to both control and isolate variables for assessing outcomes, but that these simplified texts often create as many artifacts and spurious effects in findings as (hypothetically) real world text. Ausubel’s (1976) distinctions (echoed by many others) between meaningful and non-meaningful learning and instructional text and natural real world learning situations versus artificially contrived experimental ones is extremely important here, and in this line of research and theorizing for a number of reasons. First, as Ausubel continually pointed out, the vast majority of the knowledge we learn is verbally mediated and this verbal mediation is thick and rich text supplemented by other informational aids (i.e., diagrams, charts, tables, maps and so on in this context). Next to compare a “full-blown” and “soupied-up” visual aid (diagram, graph, map) to an impoverished and stunted “(over) simplified accompanying text” (thin and poor prose) is to create one of the major “Aunt Sally” or “straw person” experimental comparisons of the past fifty years. Is it, then, any wonder that “visuals do good” and get some (weak) effects and “text does poorly” and only occasionally gets effects or adds to the effects of visual aids? The real surprise in all of this research is that impoverished text does so well and does so well so often, which strongly shows how powerful “text” is, for all of the reasons given at the beginning of this article, and how strong the effects would be if thoughtful and experimentally comparable forms of text were used in these studies, particularly in combination with visual aids of high quality.

In none of these studies, are “control visual aids” used, which would be impoverished and stunted versions of the “target full-blown instructional aid” (see Carifio & Lanza, 1992 for details). These would also be falsification conditions, and if lesser effects were not observed for these control visual aids, as compared to the complete or high quality experimental version, there would be significant problems with the theory on which the study was based, as well as with interpreting the results of the experiment itself relative to “confirmatory claims.” Few studies in science education in the last 30 years have incorporated falsification into their designs, which is the essence of modern science and modern experimental methods, which is not only a major and grievous flaw, but also belies the claims that research in science education is becoming more theory driven (see Lawson, 2007). But all of these problems are not just true of research in science education. However, “visual aids” in science and the actual practice of science are typically about depicting empirical or experimental data, as opposed to depicting conceptual structures and concept maps, and probably by a large order of magnitude, so it would seem there should be a plethora of studies on the visual depiction of empirical and experimental data and results. Unfortunately, there are few such studies in the area of science education.

In their study of the “mathematization of experience” in an eighth-grade open inquiry environment, Roth & Bowen (1994) explored the role of mathematical representations in three general science classes where students learned about the relation between biological and physical variables in the environment. Numerous and diverse sources of data (such as student and teacher field notes and journals, videotapes of student activities and teacher interviews) were examined
and coded. In general, the results showed that students increasingly used representations such as graphs, maps, diagrams and equations to support data and that the use of such representations facilitated the collaborative construction of meaning. The findings in this study also suggest that when students have tasks that require that they actively find simplifying to causal relationships in large amounts of information (data or text) they employ or begin to employ visual modes of representations to unravel and establish the relationships and relational system. Part of the problem and goal of instruction in this area is helping students to develop the habits of mind to do this type of visual modeling fairly automatically rather than only when they are (externally) required or encouraged to do so. In computer science, this “shift in behaviors” is addressed under the concept of the “management of complexity.”

As computer programs become more complex, and particularly when the phenomena they represent and exemplify become more complex (such as weather, biological systems or artificial intelligence), visual macro representation and organization techniques are best in managing this complexity, especially when they are accompanied by thick and rich micro textual documentation of all elements and relationships. “Bridging files” and elements that index the visual and textual representations are the key to effective and efficient processing, program development and program correction, particularly by self-correcting processes. A similar manifestation of this general differentiation and multiple encoding principle seems to hold for human memory and human information processing as well. As previously stated, visual representations and textual (semantic) representations are analogous in many ways to “quarks;” namely, they are fundamental and complimentary pair-units of a complete whole, and particularly so in science which has (and is suppose to have) very strongly coupled empirical and conceptual components. In this respect, “visuals” and “text” (in science) are quite similar to computer programs and strategies for “managing complexities” in computer science and other endeavors. Therefore, from the fundamentally unified (but differentiated and multiple encoding) framework presented here we will critically examine the research that has been done on more complex visual aids (e.g., photographs, pictures and graphs).

Photographs

Until recently, the function and structure of photographs in science textbooks has not been described. In addressing this issue, Pozzer and Roth (2003) explored the prevalence, function and structure of photographs in Brazilian high school biology textbooks. They describe representations along a continuum that proceeds from the less abstract and more detailed (e.g., photographs) to the more abstract and less detailed (e.g., equations) based on the amount of contextual detail in the background of these representations. Typically, photographs have greater contextual and background detail than graphs and equations. The complete continuum in order of increasing abstraction is as follows:

*photographs ➔ naturalistic drawings ➔ maps/diagrams ➔ graphs/tables ➔ equations*

It is interesting to note that as contextual details are lost on this continuum (i.e., the “flesh” or surface structure or topology), relationships and structure become more transparent and directly observable (i.e., the “skeleton” or deep structure).
Posner and Roth’s analysis revealed that all photographs could be grouped into four functional categories: decorative (5.4%), illustrative (35.1%), explanatory (28.4%) and complimentary (31.1%). Although a potentially valuable framework for textual representations, research is needed to validate this system and to better understand how these different functional categories influence learning, as this seminal and important research area has not proceeded past the descriptive level, except in the fields of advertising, medicine, law and espionage. A potential theoretical framework for conducting this research as well as the design characteristics of this research are outlined in more detail below. Also, the research on “moving pictures,” animation, and dynamic graphics and visuals as well as multimedia and web presentations, audios of various kinds, demonstrations, labs, lectures, blackboards and power point presentations have been purposefully excluded from this review for several reasons. First and foremost to begin to sort out and make better sense of the research and theories on “static” text and visuals and their processing, which are still the dominant mode in education, before moving on to more complex situations, concatenations and delivery systems. Second all of the problems we identify with the current research on “static” visuals and their processing also are present in these more dynamic and complex forms and often in an exponential manner. And lastly, if we can better understand, conceptualize and theorize the more elementary and easier to precisely control experimentally “static” text and visual level, it will give use a much better foundation, theory and set of experimental design principles to use in studying these more complex and dynamics forms of “text and visuals.” Given these points, we will discuss graphs next.

Graphs

Berg & Smith (1994) state: “What can be said with certainty is that graph construction and interpretation are critical to understanding and conveying information in science” (p. 529). However, research demonstrates that students and adults lack basic graph construction and interpretation skills (e.g., Shaw, Padilla & McKenzie, 1983). Prior research also suggests that students see graphs as (iconic) pictures instead of symbolic representations (or expressed relationships) of data (Kerslake as cited in Berg & Smith, 1994, p. 529). The act and skill of learning to “see,” interpret, understand and elaborate the features of a graph (or topological structure) and their relationships are an extremely difficult instructional task and something that Thom (1989), the progenitor of catastrophe theory, refers to as a process of semantic realization. Much of the research in science education as it relates to graphing aims to help students better “visualize” and understand natural phenomena (even though roughly 20% of the population cannot visual very well at all) through graphic representation (but typically without explicit focus on and development of the semantic realization, or more specifically, meaning component of the information and representation). However, the concept of visualization is not clearly defined by many researchers, which creates difficulties in comparing studies on visualizing and visualization strategies and techniques. Just because data enters through the eye does not mean that one is visualizing or that visualizing is occurring (see Kosslyn, 1994 for details). As previously mentioned, there is a range and degree to which members of the population can effectively visualize external stimuli (see Kuzendorf & Wallace, 2000; Byrnes, 2001). For example, some people can appropriately process the focal aspects of a visual stimulus even if the visual field includes distracting (extraneous) stimuli, while others do not have the ability to overcome distracting stimuli and, therefore, are unable to process the focal stimuli. The former are referred to as field-independent individuals and the latter are referred to as field dependent
individuals (see Byrnes, 2001). Research shows that field independent individuals are more likely to control extraneous variables than field dependent individuals, most probably because the former have the ability to identify focal variables or stimuli with more subtle effects (Linn, Pulos, & Gans, 1981). Although, these types of individual traits and abilities must be addressed relative to studies of students’ graphic visual abilities and problem solving strategies, few of the studies addressed in this review controlled or accounted for such individual differences.

In addressing this area of instruction, Berg and Smith (1994) assessed 7th, 9th, and 11th graders (n = 72) ability to construct and interpret line graphs using a free response clinical interview protocol, and they also compared the results of this protocol to results obtained from existing multiple-choice graph assessment tools. The specific criteria used to select the multiple-choice tests used as a comparison in this study were not described. The taxonomic levels of items and the reliability and validity results for the existing multiple-choice tests were not reported in this study either; however, reliability coefficients for the clinical interview protocol were 98.5% and 93.5%. The 72 students in this study were balanced by gender and academic ability (high or low) based on math and science grade point average, which gives a very poor and noisy estimate, particularly when combined with a “median” cutting score procedure for establishing high and low groups with no normative indications given by the authors. The interview protocol included three problems that were adapted to the interview form from existing multiple-choice graph assessment tools. Interviews were not “leading,” did not challenge the respondent, and provided no cues. No independent validations of these characteristics of the interview protocol were given. Comparisons were made between results from the interview and prior results reported from the multiple-choice tests. The interview problems/questions required students to generate and justify a line graph from textual descriptions of a situation (such as a bicycle going up and then down a hill) on a prelabeled/scaled graph. In contrast to the student-generated line graphs made during the interview, the multiple-choice results from the prior studies asked students to select from a number of existing line graphs. In this research model, however, no contextual features or differences were addressed within or between the two response modes (student generated or multiple choice), nor was the developmental level of the learner assessed.

Prior research in mathematics misconceptions has suggested that mathematical misconceptions are often a product of contextual (non-structural/non-operational) variations in the problems and naïve cognitive operations of the learners (see Nasser & Carifio, 1993). Therefore, it is unclear (and psychometrically questionable) as to whether these two response formats were measuring the same constructs or variables. Further, there are few who would argue that a multiple choice item is a “thick, rich, and full” textual representation, an experimental (and conceptual) problem addressed in detail in Part I of this review, or that a multiple choice item that had fully developed graphic answers was being combined with “thick, rich, and full” text for purposes of comparisons.

For all three problems in the Berg and Smith (1994) study, consistent discrepancies between methods were found. For example, 33% of the interview subjects drew a graph that was consistent with the “correct” multiple-choice response for problem number one. For the same problem, 38% of subjects in the interview cohort drew a graph that was consistent with an “incorrect” multiple-choice test response (which means that 29% of the subjects drew a graph
that was different from the correct and incorrect response given on the multiple choice problem, an outcome that was apparently ignored by these researchers. However, almost all of the subjects from this 38% provided an explanation during the interview that deemed their answer to be correct. Again, nothing was said about the correctness of the explanations of the other 29% of the students whose graphs were different from the correct and incorrect multiple choice responses. Based on these and similar findings, Berg & Smith (1994) note that the multiple-choice graphing assessments may not encourage more than a superficial approach to the problem. They also caution that: “Multiple-choice formats do not allow ‘student-generated’ responses or reasons. When students are not given the opportunity to provide explanations, researchers and teachers may make invalid assumptions about student ability to construct and interpret graphs” (p. 538), and that “when students draw their own graphs or verbalize their own responses (instead of choosing) the results can be much different” (p. 539). Similar assessment findings and conclusions were reported by Lederman and O’Malley (1990) who observed that many high school students understanding of the nature of science were judged by researchers to be incorrect on the basis of pencil and paper questionnaire response items, but when given a chance to verbalize their views during post-questionnaire (follow-up) interviews, student answers and explanations were deemed to be consistent with the defined correct interpretations.

The above findings are certainly expected from a cognitive learning theory perspective, since student understanding and assessments should improve with increased justification, dialogue, elaboration and probing. What is a bit more questionable is whether comparing a free-recall interview and multiple-choice tests, as used by Berg and Smith (1994) and Lederman and O’Malley (1990), is a valid comparison in the first place. For example, each of these response formats would appear to require different cognitive strategies and problem solving heuristics, and both are likely to have very different contextual features that may or may not be measuring the same constructs that interact with the learner. Next, the interview protocol is more like the interactive “guide discovery” procedure (as compared to the non-interactive multiple choice format) that Van Meter (2001) found to be superior in his study of visual aids, and for the same underlying reasons (e.g., cueing, guided arousal of long term semantic memory schemas and knowledge structures, response elaboration effects, and non-verbal performance feedback). Lastly, this study again also underscores that the skills for construction and interpretations of visual representations are not well and routinely taught (like reading) from an early age so that they are automatic responses (and so taught across different testing formats), and that this is part of the problem in terms of variables that need to be controlled in both research and instruction. Somewhat parenthetically, it should be noted that the problem of not systematically teaching students to read and understand graphs (fairly automatically) is as much a problem of reading and language arts curricula with their excessive text dominance as it is a problem of the science education curricula. Without doubt, graphic representation reading skills as well as graphic representation construction skills should be “taught across the curriculum,” as they are that important now and in the coming world.

In a follow-up experiment reported in the same study summarized above, Berg & Smith (1994) studied 1,416 subjects to further assess differences between student-generated or free recall (F-R) assessments and multiple-choice (M-C) assessments. The subjects represented and equal number of males and females and included high, medium and low ability groups (“ability” was not defined). Using the same three problems used in the first study, students completed
either an M-C assessment or an F-R (interview) assessment. The results indicated that the type of instrument used influenced the rate of correct responses for two of the three problems in favor of the F-R method, but only one of these differences was found to be significant at the .05 level. For each of the three problems, males had significantly more correct responses than females and high ability students provided significantly more correct responses than the medium and low ability groups. This gender difference in favor of males is consistent with results reported on standardized aptitude tests, where males significantly outperform females on mathematics subtests that involve the interpretation of quantitative data in diagrams, tables and maps (Aberg-Bengtsson, 1999). Gender is a very weak surrogate also for field independence-dependence measures and visualization and imaging measures, but in often contradictory ways as females tends to be slightly better at imaging than males, which is different from visualization (see Ashen, 1992, Kosslyn, 1994; Kuzendorf & Wallace, 2000). Therefore, nothing really “substitutes for” direct measurement of these important individual differences when it comes to researching or speaking about the effects of “visual aids and/or text,” as the effects simply are not homogeneous and are quite often disordinal interactions (see Solomon, 1971; Cronbach & Snow, 1972; and Kuzendorf & Wallace, 2000 for details).

The type of instrument used in the Berg & Smith (1994) follow-up study did not significantly influence the results by gender (i.e., there was no interaction). Instrument type did have a direct effect on the frequency of responses that represented a “picture-of-the-event-diagram” (a graph that pictorial and superficially resembles the problem, but is not the correct response). In each of the three problems, subjects in the M-C group gave more picture-of-the-event responses and this was significant at the .001 level for each problem. Based on these findings, the authors contend that graph assessment tools need to focus on student-generated graphs as well as student justification of the graphs constructed. The results also indicate that student-generated graphs (versus mere selection of a graph) are more likely to encourage cognitive engagement and are necessary to understand student ability in this domain of instruction. However it should be pointed out that anything that requires or assists a student to process more information at a greater level of depth or for longer periods of time (the F-R instructional strategy does both) will enhance learning, and that this type of enhancement is not restricted to “visuals” (such as diagrams and graphs). However, realizing that any student or information processor must “re-contextualize” a visual representation so that it is “thick and rich” again with details and meaning is a needed skill for the student, and this re-contextualizing process is where visuals and text become “one” and “propositionally equivalent” in working memory without doubt for subsequent and further processing. These points are important points for researchers and teachers to understand to disentangle a number of complex things about “visual aids” questions and views, as well as exactly what “student justifications” of graphs they have constructed are instances of as cognitive productions, and what specific generic cognitive skill they are actually invoking (and teaching) when the request for justifications is made or required. The power of reconstructive elaborations (the generic cognitive skill) of any kind by students to enhance learning and higher order learning in particular is well documented in the literature (see Scruggs & Mastropieri, 1989), as is the power of meta-cognitive skills and self-regulated learning (see Corno, 1986; Pajares, 1996; and Paris & Paris, 2001 for details). The power of “guided and directed processing” has already been discussed, as well as the power of long term memory activation and arousal, arousals and integrations of different representations of information (a variation of the encoding variability effect), depth of processing and amount of
time spent processing including reflection and reflective review. All of these general factors enhance learning, but the key is to understand that visual aids and interactive instructional processes are just two ways or strategies for activating these processes in instruction. However, one of the critical things that needs to be done in this area of research is to both describe and evaluate visual aids in terms of the degrees to which they activate and maximize these general processes, while the other key activity is to assess which components of visual aids activate and maximize which of these processes.

As noted in the previous two sections of this article, an important and recurrent instructional theme relative to illustrations, diagrams and graphs is the link between student-generated products and a better understanding of the problems at hand. However, this is not strictly a student-based phenomenon, but has also been noted among practicing scientists and engineers. Roth & Bowen (2003) explored this issue to some degree in their study of graphing expertise, which analyzed the interpretation activities of scientists related to familiar and unfamiliar graphs and their associated text. In this study, 16 practicing scientists (7 Masters and 9 Ph.D. prepared) including 13 ecologists, two physicists and one forest engineer were asked to interpret three graphs commonly found in undergraduate ecology textbooks and trade journals and one graph from the experts personal research. The analytical data was generated from videotapes, transcripts and other artifacts generated by the respondents while they attempted to interpret and make sense of the graphs.

Results indicated that the experts had little difficulty interpreting, understanding and elaborating the graphs they were familiar with (i.e., their own graphs), but that their reading and interpretation of the personally “unfamiliar” (but not professionally foreign) graphs was generally a more labored, complex and iterative process. This finding concretely illustrates more than any other the point made at the beginning of this article about the importance of “accompanying thick and rich text” for any graphic or visual aid to be truly effective, and where the true effect of the graphic or visual aid is coming from in a particular context. With the “familiar graphs,” these scientists provided their own thick and rich accompanying text from their long term memory stores of both abstract knowledge (semantic long term memory) and professional experience (episodic long term memory). With the “personally unfamiliar (but not professionally foreign)” graphs, there was no “accompanying text’ explicitly provided and these scientists had no implicit accompanying text that they could provide or generate and utilize from their long term memory stores. So any effect of any graphic or visual aid is due to the objective properties and characteristics of the graphic or aid and the particular characteristics and properties of the person processing the graphic or aid and the two cannot really be separated and must be specified in any theory or experimental situation. Interestingly, many of the difficulties the experts encountered with unfamiliar graphs are similar to those difficulties reported by students, as previously discussed. For example, some experts failed to identify the salient features of the graphs or the meaning of the graphs and their usefulness in the “real” world. These findings underscore the previous point of graph reading, interpreting and elaborating skills being acquired skills that become automatic, but tend not to be taught or learned in any broad, generic and systematic way (i.e., visual representation literacy), and thus the difficulty with unfamiliar graphic representations and forms.
It is important to point out that the study by Roth and Bowen assumes that professional scientists are a homogeneous unit of analysis with respect to reasoning level and other factors such as field dependence that have been shown to significantly influence and interact with the interpretation of graphic material and other visual-spatial reasoning tasks (see Linn, 1978; Linn, Pulos & Gans, 1981). Empirical studies related directly to middle and high school student graphing ability also demonstrates that formal (propositional) reasoners significantly and consistently outperform concrete and transitional reasoners in graphing activities and assessments (Ates & Stevens, 2003; Berg & Phillips, 1994; Wavering, 1989). From a neo-Piagetian research perspective, these types of cognitive issues are important to consider in any study that looks to understand the performance of learners (expert or novice). Unfortunately, many of these individual differences are not carefully addressed, conceptualized or controlled for in the studies reviewed above.

Summary

Visualization is a cognitive and adaptive strategy in the “management of complexity,” particularly when, in real time, that complexity is not linear and/or does not occur linearly, and the “plane” or topology of the streaming “surface flow” must be “broken” (differentiated) into additional dimensions and surfaces for parallel processing. Visualization (and its cousin mental imaging) is a cognitive strategy and ability for making the “streaming text” of reality multidimensional, cognitively, and efficiently organized for purposeful and effective action interactively with semantic and episodic long term memory stores, which are text (propositionally and networked) based according to the consensus of current main-stream cognitive theories of processing, human (and machine) behavior and learning. It should be clearly noted that Piaget and the research of others on cognitive development has clearly shown that a central feature of the qualitative increase in cognitive capacity and capabilities of the developing person is the number of dimensions in processing that the person can attend to and handle and attend to and handle in a coordinated fashion, as every person’s dimensional and informational processing capacity and capabilities are always limited (see Byrnes, 2001 for details). Thus the central importance of George Miller’s (1956) famous cognitive heuristics of 7 +/- 2 elements and their relationship for adult processors, as the heuristic is less in young adults and children, and the need to assess the level of cognitive development of subjects in all studies on visual aids and learning, as well as the differential aptitudes and abilities of subjects (see below for details).

Visualization can be seen as a highly developed form of eidetic memory (i.e., the ability to accurately recall and manipulate images), which is one of several highly specialized memory stores (or capabilities) that are somewhat like “specialized co-processors” or additional information and memory representational capabilities to the “main” and central processes of episodic and semantic long term memory. Eidetic memory tends to be strong and most prevalent in children, and it tends to become less strong, accurate or prevalent with the onset and development of language and semantic forms of representation, until it is a fairly rare gift in its strong and highly accurate form in adults (see Stomeyer & Postka, 1970; Minsky 1986; and Searleman, 2007 for details), which is additional empirical and developmental supporting evidence for the unified text-visuals propositional representation model presented and elaborated in this article. Eidetic memory also provides the beginning of a theoretical linkage to
photographs (and similar thick and rich images) and their processing and effectiveness (or lack thereof). Photographs are the type of visual about which the least is known instructionally and the least amount of instructional research has been done relative to the teaching and learning of science content. Episodic long term memory, on the other hand, tends to be “story based” and “linear,” whereas semantic long term memory tends to be hierarchically “organized” and “spatial” in the sense of the “geometry” of a dynamic network of propositions and information processing acts and cycles. The “spatial” characteristics of semantic long term memory are “visual” in nature in the sense of the “organization” and “geometry” of the dynamic network of elements and relationships (i.e., propositions) that are a person’s semantic (core “academic” knowledge) long term memory store (see Carifio, 2005 for further details).

Other specialized memory stores are “photographic memory,” which is perfect recall of text (a very key point), and “bubble” memory (see Winograd & Neisser, 2006). Bubble memory is the isolation and isolated storage of currently disassociated information that is not integrated, or only very partially integrated, into the long term memory (propositional) network, such as information related to trauma and very frightening experiences (typically one trial learning events) and information that is highly emotionally charged and has been “disassociatively learned.” Bubble memory is one of several alternative explanations of and counters to the separate coding and different language (mental models) theories and views of visuals and visual processing, which are essentially behaviorist in both origin and character and dominate educational research and educational views in this area currently (see Clark & Paivio, 1991; Schnotz, 1993, and Meyer, 1993 for explication of these views). Bubble memory is also related to “flash bulb” memory, which is the uncontrolled and rapid rush of highly detailed and highly charged images, emotions, sounds, smells, and sensations of a “life defining moment” or traumatic experience. Flash bulb memories tend to “stun” and “immobilize” a person and tend to be highly “pictorial” and “photographic,” and as close as one gets to actually reliving a prior experience in its un-interpreted and unprocessed form (see Winograd & Neisser, 2006 for details). In their very weak form, they can be the recall of the “flash of insight or genius” or a ‘first kiss’ or similar experience, although these recalls tend to be “highly elaborated” from long term memory stores and capabilities, and a source of both distorted and false memories (see Loftus, 1997 for details).

Flash bulb (and bubble) memories are “visuals” and “visual aids” (and particularly so now with virtual reality learning environments), but they are not the kinds of “visuals” or “visual aids” that are part of the view of these phenomena that one finds in the science education literature or studies typically done in education about visuals, visualization, imaging and their effects on learning. Why is this point important? One reason is that a popular research methodology and technique in this area is to have students generate diagrams, visuals of various kinds, and graphs (because this behavior is the “constructivist” thing to do and study) on the assumption that this generative process is a straight forward cognitive (and, of course, un-emotional) process with no “intrusions” from any other sources at all (see Loftus, 1997 and Minsky, 2006 for a very different view of the nature of “generative” and “constructivist” processes and how un-straight forward and “intrusion prone” these processes are). Recall the unreported 29% of student generated responses in the Berg & Smith (1994) study that matched no correct or incorrect answer at all in their array of possible answers. How could such responses have occurred (and with such a high frequency) if there were no “intrusions” into the
generation process at all, as implicitly assumed by researchers in this area and those using this visual generation technique? Intrusions are a major problem in all generational and constructivist processes (see the classic work by Bartlett, 1932 on this point, and then Neisser, 1967 for details), as well as in the processing of and reactions to “ambiguous” or de-contextualized stimuli, which all of these visuals and visual aids are to some degree in the studies reviewed here (see Herman, 1992; and French & Harris, 1999 for details). Having a fuller and more complete theory of “visuals, visualization, visualizing, and cognitive and affective processes” including intrusions and intrusional processes, which much of Vygotsky’s, Piaget’s and Freud’s work on thinking and cognition is about actually, is very important to designing, analyzing and interpreting studies in this area and on this topic.

The points of importance here are that neither visualization, visuals, or the human memory systems associated with them and with processing them are well conceptualized, well theorized, or up-to-date in any of the studies of visuals, visual aids, visual learning, or “school instruction” (text plus visuals) we have read in the area of science education. Similar problems exist with the concept of “text,” or the relationships between “visuals” and “text,” and the elementary fact acknowledged that all visuals are “situated visuals” and typically situated in text or a given context or implicit or explicit set of contextual information. This basic and core problem is also one of the chief reasons why the questions posed in these studies are weak and overly functional, and the treatments are so amorphous and unspecified, and why the experimental designs have so many flaws and so many uncontrolled variables, as well as why visuals with impoverished texts really do not tell us that much or that much that is new or we do not already know from the study of ambiguous and de-contextualized stimuli. This area in science education is and has not been “theory-driven” in the real and significant sense of this term, which is why the findings to date on this very important question and topic in science education have been so amorphously functional and contradictory.

No study we could find used control visuals or visual aids or other forms of falsification in the study’s design. No study sought to measure the elementary level of “complexity” of the visual aids and text used in the study, so that apples could be compared with apples and so on, and one could assess the degree to which the visual aid and/or text was close to or taxing the limited informational processing capacity of the study’s subjects or not. No study directly included level of cognitive development of the subjects or assessed the visualization and imaging abilities of the subjects for analysis and study comparison purposes, and all blithely assumed that visualization and imaging abilities were homogenous and fairly constant for all people and subjects in their study, when in fact roughly 20% of people cannot visualize or mentally image very well at all (Ashen, 1992; Kosslyn, 1994; Kuzendorf & Wallace, 2000). Given these basic empirical and scientific facts, what homogeneous effects would one expect to find in an experiment on visuals or visual aids, and why would one not be assessing the interaction effects between subject type and visual aid type and characteristics, particularly given the findings of Roth & Bowen (2003) about the inabilities of practicing adult scientists to process graphs that were not professionally foreign to them but rather just unfamiliar to them? There is little doubt from the current basic research literature that the “visual aid” question is largely an “aptitude-treatment” interaction (ATI) question and findings to support this point are actually included in Cronbach and Snow’s (1972) classic book on ATI’s and ATI models. Why then was not one authentic ATI design found in any of the current studies of “visuals” in the science education
literature, or in any of the literature cited by these studies? Our view is because research in science education for the past 15 or so years has been conducted in somewhat of an insular and parochial bubble, if not vacuum, without a great deal of connection to mainstream literature and theory, as such connections and connectedness is no longer fashionable, or as sought as it was in the past. Obviously, this state of affairs is one that needs to undergo some rapid change in our view.

Conclusions

This article addressed a number of issues, points and concerns relative to some of the empirical research in science education and related fields that informs our understanding of the factors that influence instruction involving visual materials including diagrams, illustrations and graphs (DIGs) and their logical instructional derivatives and correlates. To summarize this critical review, eight key points and propositions (or conclusions) are listed below that are somewhat more concrete and specific than the general conclusions given above. These eight points are as follows:

1. The role and importance of text structure (as well as “stunted” text) is not addressed (or controlled for) in any of the studies we have read, but particularly those studies that have students use text to generate graphic “visual” representations, which are used to measure student understanding. The importance of text structure relative to processing and learning outcomes is well documented by vanDijk (1980), Kintsch (1986), and Britton and Graesser (1996), but particularly so by Frase (1969), Shavelson (1971) and several other studies on the conceptual presentation order of the text and its influence on the resulting concept maps students have and produce after reading the text in question. Further, the implicit view of text and text processing in many of these studies is overly simplistic and assumes that text and language are strings of words that are fundamentally different to “visual” representational systems such as diagrams and graphs, which are seen as being basically no more than “pictures.” These views are diametrically opposed to the mainstream research in cognitive and computer science and the area of artificial intelligence which recognizes that verbal, mathematical and visual systems (and displays) are all forms of representing (deep structure) information, and all have their own but similar languages. Further, modern cognitive psychology and cognitive sciences contend (and have vividly demonstrated) that all visual information and stimuli may be represented propositionally (as with computers) and at different levels of depth. It is the propositions that define the relationships in a (deep) knowledge structure and/or (surface) display (or topology) and provide the semantic (meaning and interpretation) component in any system (Penrose, 1991). This propositional representational view is in fact the central and core postulate and contention of modern cognitive learning theory (Ashcraft, 2002). The point being missed in this research is that propositions define the relationships and provide the semantic component in any system, but particularly so for visuals and visual aids where the semantic component is often missing, implicit, or assumed to be in the long term memory stores of the person processing the visual (all leading to failed learning outcomes typically). The basic theoretical and empirical fact is that “visuals” and “text” are “quarks” and inseparable components of a fundamental informational unit (much like the multiple files of a webpage that appears as a unitary display or the base and G cleft in a piece of music as one complex sound), with each of the components
included coding the knowledge kernels in different but complimentary ways to create a “thicker, richer, and more connected and integrated” representation. The current conceptual “dualism” between text and image, then, is what must be changed for research to progress in this area and on this topic.

2. In general, diagrams, illustrations and graphs are not well defined or validated instructional tools and have no clear underlying theory as to what they are. This point is relative to independent panel review of the visuals to see if they are judged to have the characteristics and features claimed (see Lanza & Carifio, 1991 for details), as well as use of control visuals (see Carifio & Lanza, 1992) and more objectively describing visuals in terms of their “complexity,” even if those objective descriptions are relatively elementary.

3. Individual differences that have been linked to an individual’s ability to visualize—such as gender, developmental (reasoning) level, field dependency, prior knowledge, and basic visualization and imaging aptitude—were not consistently or carefully addressed or controlled for in the studies reviewed in this article. These factors need to be measured and/or controlled for and a basic aptitude-treatment study design used, particularly as the odds are that the “visuals” will be relatively ineffective for roughly 20% of the subjects in the study, which means that effects at the aggregated level of analyses typically done in these studies will be significantly under-estimated for subjects with good visualization ability. This factor is most probably the cause of all of the close but not statistically significant findings in so many of the studies reviewed in this article.

4. None of the studies examined here explicitly addresses the central tenant of cognitive psychology that the human information processor is severely limited and uses various strategies to transcend this limitation. With respect to visuals and text, these strategies may involve the learner moving back and forth from the text to the visual (and vice-versa) in a dynamic, chaotic and unpredictable fashion. Instead, many of the studies examined above adopt a neo-behaviorist view of learning and memory that assumes that instructional materials have an incremental, “additive” and “linear” effect on learning and that insight, restructuring and reorganizing, self-managed meta-cognitive processes and global changes do not occur and are not self-initiated by the learner, which they most certainly are for many if not the majority of learners (see Carifio, 1978; Paris & Paris, 2001). This point is important as some of the effects found in some of these studies could be due to the lock-step programmed instruction-like learning and testing processes used that restricted learners from actually behaving more “naturalistically.” Text and visuals are two different “interplaying systems” or instructional components (each with strengths and weaknesses) and students move back and forth between the two according to the beat of their own drummer to learn and understand in iterative cycles. This process was severely restricted in several studies which could have, in part, accounted for some of the results observed. As the Van Meter (2001) study found, student-guiding processes were more effective in promoting learning than student controlling processes and particularly for more self-directed learners.
5. High quality (and valid) diagrams, illustrations and graphs (DIGs) supported by appropriate (supportive) explanatory text appears to provide cognitive guides, aides, referents and cues that direct attention to the salient features of the DIGs. How differentiated this effect is among different learner profiles is not known and needs to be established.

6. Some learners can and do construct DIGs to structure and organize complex ideas, concepts and information when directed to do so or spontaneously to address novel problems. Who these learners are as a subgroup is not known relative to DIGs and needs to be established as do the procedures for inducing these behaviors in other students. There are, however, some strong suggestions relative to answers to these questions (see Pajares, 1996; Paris & Paris, 2001).

7. The studies reviewed here seem to suggest that domain-specific low-level (reasoning and experience) learners may have more to gain by using DIGs than domain-specific high-level learners since the former are less likely to spontaneously construct such representations. Much better designed studies and more sophisticated analyses are needed to both clarify and confirm this suggestion and the possibility that it may be confined to what Bruner called iconic (visual) representations and processes and lower levels of cognitive development, or are more generalized as these researchers implicitly contend.

8. Graphing expertise and ability may not transfer to other domain-related concepts or unfamiliar applications, even for adult experts. This may be due to “visual transfer” being more local or the “near” form of transfer than suspected or it may be due to not having conceptualized the “deep structures” of graphs and other visual forms and explicitly taught these “deep structures” and how they broadly generalize across types. The weight of indirect and direct evidence seems to suggest that the later may be the case as formal (operations) reasoners outperform transitional and concrete reasoners on various DIG activities and assessments in both passive and active (i.e., self-generating) forms.

None of the above points should be interpreted as mutually exclusive factors influencing performance, as learning to use and understand DIGs meaningfully is a dynamic process involving a number of interacting factors such as the quality and structure of the associated instructional text, gender, prior knowledge, reasoning ability, level of field dependence, prior instruction and familiarity with the specific DIG and its application to specific problems.

One of the main purposes of this two-part article was to review the literature on learning scientific-technical content from “static” text and visual representations that is available in the current science education and related literature. This review was carried out in part because such a critical review was needed and in part to establish a baseline on what is known about students learning formal models (both scientific and non-scientific) which tend to be visually depicted today for a variety of reasons including the complexity and detail of our current knowledge about a broad array of phenomena including a wide range of non-linear and dynamic phenomena. This later point is one of the reasons why our focus on a partially quantitative description of some
kind of the complexity of the visuals being studied was so strongly emphasized relative to the
studies reviewed. Another reason for this emphasis is that formal models and the necessity of
learning formal models in “some intellectually honest way,” as Bruner (1965) used to say, even
at the elementary school level is only going to increase and increase quickly and markedly, and
not just in science education. We know exceedingly little about students’ abilities to learn
formal models or the procedures that facilitate such learning. Such a deficiency currently is
more than a grave one. This deficiency is one of the major limiting factors on the acquisition of
scientific literacy and science knowledge now and in the coming decades. The research studies
that have been done in science education on “visuals” and “visual aids” are but faint hints and
whispers of selected aspects of students abilities to learn formal models and what procedures
facilitate such learning, and this is one very important reason why such studies in science
education must be better conceptualized and executed, as well as why the numerous flaws,
problems and difficulties we have identified are so important. But science education and science
education research must also go far beyond the remediation plan we have outlined in this article.
Both must become “rapidly expansionist,” empirically and theoretically, and truly
interdisciplinary in character to address the problems of how students (best) learn formal models
and to think using them and to make new ones. This challenge is one of the challenges in
science education currently, as well as other numerous disciplinary areas.

To help facilitate this process, we have tried to connect this literature (and science
education) to other literature, theory, and empirical findings in a number of disciplines
to initiate the forging of this more general and eventually standard framework for theoretically
and empirically investigating the “visual aids” question, but more importantly the far more
inclusive acquisition of the formal models question. With respect to both of these questions, the
work that has been done on critical feature theories of perception and representations (Neisser,
1967; Norman, 1981) that were carried over to concept learning, schema development and
visual/pictorial scene analysis and understanding by computer vision research programs that was
initiated at MIT in the mid 1980s (Minsky, 1986; Minsky & Papert, 1988) and has been on-going
since would be good sources to examine hypotheses and for validated models and theories, as
well as to rule out rival models and theories. The on-going work of Kossyln (1994) and others
on “brain and image” and visualization and visual and envisioning processes should be read and
followed closely, as well as the on-going work of Ahsen (1992) on mental imaging and
visualization which encompasses concrete enactment (doing/behaving) processes as well as
cognitive and affective processes. Ashen’s work and views about images and imaging have been
particularly effective with special needs children.

Recent work in virtual reality and the representational systems and components of these
dynamic and open-ended systems should be examined for “transferable knowledge,” particularly
those that utilize speech and text as well as computer animations of various kinds. Virtual reality
is far closer to the instructional classroom than most suspect, and is already being used with great
success therapeutically and in medicine (see Hoffman, 2004). The Artificial-Intelligence
(cognitive) based computational philosophy of science models of Thagard (1988) and the
mathematical-psychological models of decision making and choice developed by Tversky and
Kahneman (1981) should be examined for “transferable knowledge” about visuals, text, and
formal models and formal model development and revision processes. The computer analysis of
visuals as well as text have grown in leaps and bounds in the last decades with much new, very
generalized, and “transferable knowledge” available, as anyone who watches even a little television today can attest. For example, what would an AI guided “computer enhanced” graphic, diagram or visual for instructional purposes look like (and the same for text)? We need to better and more generally and theoretically understand “visuals” and “visual aids” so that we can better approach the real question and real challenge which is how students (best) acquire formal models and learn to think with them and adapt them. That is the science education and science education instruction of the future and the pressing challenge we must address.

References


Carifio, J. (2005, July). Towards a standard integrated information processing/cognitive model of


Herman, J L (1992). *Trauma and recovery: The aftermath of violence from domestic abuse to political terror*. New York, Basic Books.

Herman, J L (1992). *Trauma and recovery: The aftermath of violence from domestic abuse to political terror*. New York, Basic Books.


Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.


Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology, 93*(1), 129-140

