Psychometric Scaling: Avoiding the Pitfalls and Hazards

Peter G. Engeldrum
Imcotek
Winchester, MA 01890
pge@imcotek.com

Abstract

Psychometric scaling is widely used in the imaging filed for obtaining scale values of image quality and its attributes. Although widely used, scaling has its pitfalls and hazards. This paper reviews the process of scaling and provides some practical hints for conducting psychometric scaling studies and avoiding the most common hazards. The scaling process is described along with suggestions for sample selection, observers, task instructions, and, presenting and viewing the samples. The goal of this presentation is to provide practical guidance to practitioners of psychometric scaling leading to improved efficiency of the scaling study, and, increased precision and accuracy of image quality and attribute scales.

Introduction

Simply stated, the goal of scaling is assigning numbers to image quality and the "nesses," or attributes. A scaling study is the process of establishing these numbers or scale values. On the surface, it appears quite simple to ask observers to express a judgment or an opinion about some images, but it is subtler than that. The purpose of this paper is to give an overview of the process of designing and conducting scaling studies, and to provide practical guidance or hints. Reference (1) includes more details on the topic of psychometric scaling.

It is a practical impossibility to cover all possible factors affecting the results of a scaling task and provide specific recommendations for each of them. Attempting to do so would take a book in its own right. The objective here is to limit the hints to some key areas relevant to imaging.

The Scaling Process

At one level the scaling study appears deceptively simple. Anything so simple would not appear to need extensive planning. However, diving headlong into a scaling activity without a plan is almost guaranteed to yield poor results, results that may be useless or totally inaccurate. In a schedule-driven product development environment, this headlong dive is more the rule than the exception, though. The message is simple: before you begin, you need a scaling study plan. The following sections describe some of the many factors that need careful consideration in developing a scaling study plan.

In most applications one is interested in scaling either image quality, or the perceptual attributes comprising image quality called the "nesses." The word "ness" is used as a shorthand notation because it reflects the ending (suffix) of most words describing perceptual attributes; e.g. lightness, brightness, and colorfulness.

The process of developing the scale values for a "ness" or image quality consists of seven basic steps:

1) Select the samples (stimuli).
2) Prepare the samples for observer judgment.
3) Select observers.
4) Determine observer judgment task or question.
5) Present samples to observers for their judgment or preference.
6) Collect and record observer responses.
7) Analyze observer response data to generate the scale values.

These steps interact, often in unforeseen and unpredictable ways. Serious consideration and planning of the scaling study is needed for successful results. The remainder of the paper discusses issues and provides some useful suggestions for completing the first four of the seven steps outlined above.

Sample (Stimuli) Selection

Selection of image samples, or stimuli, is governed by the objective of the scaling study and many other practical factors. Sample selection is, in practice, one of the most difficult parts of the scaling study, and is not often given the serious attention it requires. Failure to collect or generate a suitable sample set has derailed numerous scaling studies because many of the selection factors are elusive or undefined. By focusing on a few critical factors, the process of sample selection can be substantially simplified.

Four key factors that need to be addressed during the sample selection or sample generation phase of the scaling study planning are:

1) What categories should the samples represent?

2) What range and distribution of "nesses" should the sample set contain?

3) What image size should be used?

4) What image content or image elements should the samples contain?

Although all of these factors are key, there is no universal optimum set of these four elements suitable for all scaling studies. The choice of elements will depend on many practical considerations and the necessary trade-offs.

Categories

Bartleson[2] proposed five categories to describe samples of imagery, and listed their basic properties. The real value of Bartleson's categorization is that it represents an organized way to make a rational sample selection; and, conversely, identify what properties the sample does not possess.

Random and Independent-Random and independent sampling of images, although statistically interesting, is difficult in practice to implement. A major issue is the difficulty of defining a population that can be randomly and independently sampled. There is no single reservoir of "the population" of images, although, as digital imagery continues to develop, there are an increasing number of firms on the Internet offering image files in an increasingly large number of image classes.

Stratified-Stratified sampling or imagery is becoming more practical due to the wide availability and accessibility of image databases. Imagery can be defined in classes such as text, graphics, and photographs. The class "photographs" can be further stratified into various subclasses such as landscapes, portraits, and groups of people. Stratified sampling is a very practical approach with wide availability of digital image files.

Contrast-Contrast sampling is common in a product development environment. Usually there is an interest in knowing the quality requirements or performance of a particular imaging device with respect to some class of imagery. The market for the product shapes the classes. Selecting imagery classes relevant to the product application is also efficient because it ignores irrelevant classes.

Purposeful-A purposeful sampling can be extremely useful during product design. During the product development process, questions arise that require engineering trade-offs. Often the prototype product produces some unexpected "ness," and raises the question, "What level of the (unwanted) 'ness' is acceptable?" A set of sample images that exhibit various levels of the "ness" in question would comprise the sample set to be used in a scaling study.

Incidental-Incidental sampling is, arguably, the most widely used sample category in imaging product development. Typically, a set of images is selected as the "reference" set, supposedly representing product performance requirements. These images then become the "gold standards" or "sacred samples." These incidental samples are often selected by the product development team to represent a readily understood image quality contract between the relevant product development organizations-marketing and engineering, for example.

Range and Distribution of "Nesses"

The samples define the context of the scaled "ness" or image quality. There are two aspects to this depending on whether a "ness" or image quality is being scaled. When scaling a "ness," if the "ness" of interest does not vary in the sample set, any resulting scale cannot be a measure of the "ness" in question. The context was incorrect and it will be a scale of some, perhaps unknown, "ness." In scaling image quality, the context is the specific set of "nesses" and their range in the sample set. It is common in image quality scaling studies that the quality judgment varies due to the variation of only one "ness." The resulting scale from this sample set will not be one of image quality, although it may be labeled as such, but a scale of the single "ness" that varies in the sample set. Much care is needed in identifying the "nesses" in a sample set in order to avoid these pitfalls.

With some scaling methods, the distribution of the "ness" in the sample set can also have a significant influence on the scale values. In category scaling, observers have a tendency to use all categories equally often[3]. For example, if a large fraction of the samples have high values of a "ness" and only a few have low values, observers tend to make fine discrimination at the high end and lump the low-valued "nesses" together in the bottom categories. This judgment behavior results in scale distortion. The best solution is to have equal numbers of samples that uniformly span the range of "ness" of interest.

Generating samples using computer image simulation or rendering techniques can help achieve the required "ness" range and distribution.

Securing a set of samples that have the desired "ness" or "nesses" with the desired range and distribution is often a very difficult problem to overcome. It is more than worthwhile to expend the effort to select or generate a sample set that meets the scaling study requirements.

Image Size and Spatial Sampling

Image size and object sizes in the image are well-known factors in "ness" and image quality judgments[4,5].

In general, expect image size to be a factor influencing observer judgments, one way or another. The simplest strategy is to keep the size of the images in a scaling study constant, thus eliminating or minimizing image size as a judgment factor. Keeping sample image size constant does not eliminate any context-of-scene dependence factor, though.

Image size also enters in an indirect way. The scale values of samples from an imaging system should not be limited by the inherent quality of the input image or object. Evaluating the quality of computer printers, for example, should not depend on the spatial sampling frequency (pixels per distance) of the image file being printed. A good rule of thumb is to have the sampling frequency of the test samples be at least equal to the addressability of the output device.

A useful upper bound for spatial sampling frequency is about 10 samples or pixels per mm for images viewed at "normal" distances of about 14 inches. If the viewing distance changes, the upper bound on the maximum sampling interval scale changes in an inverse manner.

Image Content

There are a host of contexts, overt or implied, within which a sample is viewed and judged. A helpful rule to remember is, "Preferences do not occur in a vacuum, they are always formed relative to a context"[6].

The spatial configurations of the elements and the object content in sample images are well known "context" factors. The breadth of spatial configurations can vary from large areas of color (or dark areas), to areas that change rapidly from point to point (so-called "busy" images). This dependence of judgments on spatial configuration, or context, is called scene dependence. The term spatial configuration is preferred to scene dependence because it more accurately relates to the judgment factor.

To generate useful scales, a balance must be achieved between the spatial elements and objects in sample images. This is the driving reason for using more than one image in scaling studies. The assumption is that assorted spatial configurations will average out spatial-configuration effects.

Most people have fairly consistent preferences for a few "critical" colors, such as flesh tones, green grass and blue sky. It is well known that the preference for reproduced versions of these colors is quite different from the preference for the actual colors themselves[7]. These colors provide a context for observer judgments of image samples, and the judgments may be substantially altered by their inclusion.

Image classes are also known to exert an influence on judgments. For example, the sharpness of portraits of people and landscapes are judged differently[8].

Emotional involvement, or potential involvement, of the observer in the sample image or scene content is another context factor. It is no secret that sex sells, and for this reason advertising agencies use alluring women and men in product advertisements. The same idea applies in the scaling of sample images. The choice of sample, or scene, can affect the scale values in both positive and negative ways, through emotional involvement of the observers. Emotional involvement also applies to "my images" versus someone else's images. A bond or attachment to the persons or objects in the images causes altered judgments.

Although the emphasis has been on pictorial imagery, the same general rules hold for text and graphic images. For example, samples of text composed of unfamiliar typefaces (fonts) may create a foreign context for the observer, and not give a useful scale.

Sample Preparation

Once the samples have been selected, it is then appropriate to consider how these samples should be prepared for presentation to observers. Careful preparation will not only preserve the samples, but will reduce unwanted, and often unknown, influences on observer judgments.

Sample Handling and Maintenance

When the scaling study requires a large number of observers, such practical issues as routine sample handling and keeping the samples clean become important.

One useful technique for keeping the sample images clean is to mount the samples on a rigid base such as cardboard. Care needs to be taken to ensure that the color of the base or backing does not alter the appearance of the samples in undesirable ways.

Images generated by some imaging technologies are prone to damage due to mechanical abrasions of the imaging material. A cover of heavy paper or plastic material, hinged on one edge of the sample, is one means of providing protection. If the cover is transparent, it can also minimize or hide surface textures from the observer, which may be useful in some situations.

Sample Border or Mask

In addition to mounting on cardboard, placing a frame, border or mask around the samples has some advantages. The frame, border, or mask is usually a neutral gray cardboard, which serves two purposes. The first is to mask off, by covering any white border surrounding the sample. Masking the border may not be appropriate if the whiteness of the image substrate is of interest. However, if the sample set is produced on a variety of substrates, using a mask around the edge of the sample will eliminate substrate whiteness as an observer cue, thus eliminating the chance that an unwanted "ness" might influence observer judgments.

Some imaging technologies have built-in cues, such as substrate thickness, tactile feel, image gloss, and image surface texture, that let observers deduce substantial information about the sample. A simple mask or border, in conjunction with a backing material if the image substrate has low opacity, can minimize the cues and allow the observer to focus on the "ness" rather than the imaging technology.

Another important use of the gray mask is to provide a constant visual reference or adaptation point. When scaling color "nesses," a constant visual reference is important because it stabilizes the chromatic adaptation point of the observer, and forces the white point reference to be in the image sample. The border or mask surrounding an image is a primary tool in controlling the image appearance and should receive careful consideration.

Sample Labeling

Hiding the sample identification helps reduce the chances that helpful observers will "solve" the visible code and respond to the sample identification and not the "ness" of interest. Placing identifying numbers or letters on the samples in areas that are visible to the observers is generally not a good practice. In addition, one should avoid sequentially labeling samples in the order of some technology variable. The preferred labeling practice is to put the identification on the back of the sample. If it is essential that some alphanumeric identification be visible to the observer, use a non-obvious code of some sort, such as a four-digit sequence of random numbers or letters.

Bar coding of samples can speed data recording and reduce errors.

Numbers of Samples

In scaling studies, the number of samples depends on the scaling method and the time and resources available. Using many samples in an attempt to cover all the sample-selection considerations discussed so far is often impractical. Depending on the needs of the study, the number of samples can be anywhere from about three to about thirty.

Scaling method and sample quantity interact considerably. Some scaling methods are most efficient with small numbers, say less than ten, while other methods may only yield satisfactory results with larger numbers. Another consideration is physical sample size. Manipulating poster-size images is physically difficult, so the number of samples must be necessarily small if the prints are large.

Observers: "Type" and Number

A commonly held belief among newcomers to psychometric scaling is that experts see things differently, or give different scale values, than unsophisticated observers. This may or may not be true, and it depends on the scaling task. Observers who participate in scaling studies are usually eager to help, and will often use various methods to provide the "correct" answers. These factors are real and must be addressed to assure high-quality scale values.

A general discussion of observer selection is outlined in reference (9). ASTM Standard E 1499-97 (1997), which is primarily oriented to color appearance judgments, provides detailed guidelines for the selection, evaluation and training of observers.

Expert versus Average Observers

Observers who have experience in judging or evaluating images usually fall into the expert observer category. In their vocation, they may learn to make very fine distinctions of the "nesses" they experience. To a much greater degree, experts can distinguish among categories of a specific "ness": their "ness"-scale resolving power is often much greater than average or untrained observers. This may become troublesome with some scaling methods, particularly category scaling, where the trained or experienced observers distinguish among categories that average observers do not. Conversely, there are applications in product quality assurance that require fine quantization or categorization of "ness" values and the detection of small differences. Trained or expert observers are needed in this type of scaling task.

For specific "nesses" that are relatively unique to a particular imaging system (a defect for example), experienced observers may give scale values that are markedly different from average observers. When this occurs, it may be due to stimulus errors, which simply means that observers are making a judgment on a Technology Variable and not a "ness." Product development personnel are often very familiar with the Technology Variables of the imaging system, so recruiting them as observers is not generally recommended. Product development personnel tend to be more sensitive to "bad" "nesses" than average observers. If the potential observer is knowledgeable about failure modes or technology variables of the particular imaging system, and the scaling study is trying to simulate typical customer response, then such observers should not be considered for the scaling study.

If the objective is to generate a "ness" scale for average or typical customers, the safest course is to use typical customers as observers. On the other hand, if the scaling task is to scale a fundamental "ness" not specifically associated with a particular imaging device-colorfulness, for example-most human observers will respond similarly.

Observer training and task familiarity both play a role in understanding the task and the speed of executing the scaling task. Experts generally give scale values similar to average observers when scaling fundamental or basic "nesses," but they often do it faster.

The most common situation is where expert and average observers often give distinctly different responses when answering a preference question. If the scaling task requires a response to any of the following questions:

"Which sample do you prefer?"
"Which one do you like?"
"Which one is best for the xxx application?"

then it is a preference task.

Confusion about whether the judgment task is a basic "ness" or a preference is probably the origin of the myth about expert vs. average observer difference. Statistical testing is warranted if there are concerns that observer group expertise may distort scale values.

Number of Observers

The fundamental advantage of having a large number of observers is the increased precision of the estimated scale value. Using more observers decreases the error in the scale value, depending on the details of the statistics of the scale estimate. However, the general rule is that scale precision increases as the square root of the number of observers.

The number of observers to use in scaling studies is typically governed by availability. Scaling studies in the imaging arena are conducted with as few as four observers, and with as many as fifty. A recommended range is from ten to thirty for typical scaling applications. This is not intended to be a rigid rule, only a guideline subject to scale precision requirements and the experience and knowledge of the study administrator. Better estimates of observer numbers can be obtained by establishing a desired scale precision, and using the statistical relationships between numbers of observers and scale standard deviation.

Increasing observer numbers only affects the scale-value precision, or variability about its average value. There is no practical way to know the absolute accuracy of a scale value, so the choice of observer number, per se, does not affect scale accuracy.

Observer Task Instructions-It's All in the Question

Next to the sample image set, observer instructions are the most significant item that controls the context of the observers' judgments in a scaling study. Sadly, it is far too common to see the observers' task instructions get only passing consideration.

To achieve useful and meaningful results, observers need to be told what they are to do.

1) What, exactly, is the attribute they are to judge, and what is their judgment task?

2) Is there an explicit or implicit context to the scaling task?

3) What criteria or definition should they use in their judgment?

These considerations are key to any successful scaling study.

What is the Attribute and Judgment Task?

A common scaling scenario is to use the paired comparison data collection method instructing the observer to answer the question, "Which do you prefer?" by selecting one of the pair. In reporting the results, the scale is termed an "image quality" or some other "ness" scale. This is grossly incorrect because the instructions to the observer are, "Which do you prefer?" No question was asked about image quality or a "ness" preference, so the final scale is nothing more than a basic scale of preference.

The general rule is specifically to ask the observer to make a judgment on the appropriate "ness" or image quality. For an image quality scale, the appropriate instruction to the observer should be something like, "Select one of the two samples that has the highest image quality."

The judgment task instructions should be clear and should avoid complex or fuzzy ideas, technical jargon, and the use of technology variable labels.

A recommended procedure is to present a set of written task instructions to the observer to read. The scaling study administrator then asks if the observer understands the instructions. The administrator needs to be alert at this point, because over-helpful observers can use this opportunity to obtain some clues about what answers you want from them. Good practice would have the scaling administrator provide concise answers, without elaboration. This is like walking a tightrope-you want to make sure the observer understands, and yet you do not want to provide background material or explanations that may bias the observer's judgments.

What is the Context?

Observer instructions and scripts frequently set the context of the judgment in a scaling task. For example, the context of the judgment can be set by suggesting that image quality is, "The quality of images you would give to friends and family." We now perform an experiment where the observers judge image quality using paired comparisons. In a paired comparison judgment, observers then may be asked, "Select the sample that has the highest image quality." The resulting image quality scale would have a context of "images that would be given to friends and family." To say that this scale is applicable to the quality of office documents is to seriously mislabel the resultant scale.

Integrative attributes such as image quality are much more context-or application-dependent than "nesses" such as image sharpness and graininess.

Even when the question and context are carefully described to the observer, there is no guarantee that the desired results will be achieved. For example, suppose the observer follows instructions and scales a set of samples according to image quality. Yet if the sample set varies only in the "ness" dimension of, say, "textureness," the resulting scale has to be called an image quality scale; image quality is in fact the question posed to the observer. However, the scale is really a scale of "textureness" by virtue of texture being the only "ness" dimension that varies in the sample set! This is, sadly, an all too common problem.

An introductory script that describes the purpose of the scaling can also establish the context of the scaling experiment. If establishing a context for the judgment is important, then explaining to the observer the purpose for the scaling may be useful. There are mixed views about this. I am from the minimalist school that believes in giving the observer only the minimum of information that is needed to do the task. Extra information may distract the observer from the task at hand. In addition, a long verbal explanation may allow the observer to pick up extra clues in order to be helpful. Finally, an excessive question and answer session consumes precious time for the observer and the administrator. Let the observer spend time giving you answers, not vice versa.

Criteria and Definitions

Depending on the scaling objective, the "ness" may or may not be explicitly defined. There are several ways to define a "ness":

1) One can use words and define the "ness" in observer instructions.

2) One can use visual references, which are often used as anchors.

3) One can let the "ness" be defined by the observer using some internal criteria.

When an observer uses his or her own internal definition, it is of no use unless the observer somehow conveys its meaning to the study administrator.

If an explicit "ness" definition is used, it should be unambiguous and easily accessible to the observers during the scaling study. A card with the written definition can provide a handy reference. Visual references are often used as anchors in graphical rating scale experiments, but they can be used with most any other scaling method. These references are often employed where a word description would be difficult, or where they supplement a written definition.

Looking Through the Haze-Imperfect Samples

Few samples used in real-world scaling studies are perfect or defect-free, particularly in the early stages of product development. To compensate, we can ask observers to ignore scratches or dirt, or not consider image composition in their judgments. Although instructing observers to ignore certain aspects or defects of the samples is not uncommon, there is no guarantee the observers will do so.

If the attribute of interest does not interfere with, or is different from, the sample defects, then expecting that observers will respond to the "ness" of interest is reasonable. However, if the sample set has streaks and you are asking the observers to judge "bandingness," it is doubtful that, without training, observers can reliably ignore the streaks and judge only "bandingness."

Clearly, the best strategy is to use samples that do not exhibit any unwanted "ness," but this is not wholly realistic. With imperfect samples, asking the observer to ignore such unwanted "nesses" via the instructions would be prudent.

Conducting the Scaling Study

A large number of possible methods for conducting the scaling study are available. There is no universal agreement on a standard method. Scripts and pilot studies are two recommended tools that can help eliminate costly errors and improve scaling studies over the long run.

Scripts

Scripts are the written sequence of procedures, questions, or instructions to be followed by the scaling administrator. The foremost purpose of the script is to present a consistent narrative to all the observers. Most observers look, ask, and listen for cues about what the scaling administrator really wants. They search for cues because they are usually interested in the scaling process, and they want to be helpful and do a good job. Scripts are usually read aloud by the scaling administrator, word for word, as a means of enforcing the consistency of presentation. Using a consistent procedure (the script) also reduces the effects of unintended moderator bias as a factor in observer judgment.

Included in scripts are instructions to observers on how to perform the task, what criteria to use in the judgment, and pointers to external and internal references (anchors) that need to be considered. It is essential that the script should not imply or refer to criteria, definitions, or other items that can affect observer judgments.

There are several significant benefits to using a script. First is the ability to test and modify it in order to fine-tune the experimental procedure. Secondly, the well-tested script can be used multiple times, and simultaneously in multiple locations, and thus ensure that later scalings will be conducted in the same manner as earlier ones. Finally, it formally documents the complete methodology of the scaling study. Using a script is no guarantee that results will be identical, since the observers will no doubt be different, and other factors may change over time. Nevertheless, using a script will reduce the influence of factors that can affect scale consistency.

Commonly included as part of the script for the scaling administrator is a list of the environment requirements, such as a specification of the lighting, a list of equipment, any associated software, and calibration methods. All of these factors can affect the perception and judgment of "nesses" or image quality. The rigorous specification of these environmental requirements will assure identical conditions for any subsequent scalings.

All of this may seem like a lot of effort just to generate a "ness" scale. However, anyone experienced in doing physical measurements usually has a measurement procedure. In this respect, scaling or the measurement of human response is no different.

Conclusion

Psychometric scaling can be a very useful, an often an essential tool in developing and evaluating imaging systems. Its utility relies on the consideration of a variety of factors that affect the observer judgements. This paper has briefly described some of the scaling study considerations and has provided some practical advice to improve the quality of the observer judgments.

References

1. Engeldrum, Peter G. ***Psychometric Scaling: A Toolkit for Imaging System Development,*** Imcotek Press, 2000, ISBN 0-9678706-0-7
2. Bartleson, C. J., in ***Optical Radiation Measurements, Vol. 5 Visual Measurements***, C. J Bartleson & F. Grum Eds., Chapter 8, Academic Press, Orlando, FL 32887, 1984, ISBN 0-12-304905-9
3. Gescheider, George A., ***Psychophysics: The Fundamentals, 3nd Ed***., Lawrence Erlbaum Associates, Inc., 101 Industrial Avenue, Mahwah, NJ 07430, 1997. ISBN 0-80582281-X
4. Cory, G. P., M. J. Clayton, & K. N. Cupery, Scene dependence of image quality, Photo. Sci. & Eng. 27:9(1983)
5. Westerlink, Joyce H. D. M, & Jacques A. J. Roufs, Subjective image quality as a function of viewing distance, resolution and picture size, SMPTE Jour., 98:113(1989)
6. Mellers, B. J. & A. D. J. Cook, The role of task and context in preference measurement, Psychol. Sci. 7:76(1996)
7. Bartleson, C. J. & C. P. Bray, On the preferred reproduction of flesh, blue-sky and green-grass colors, Photo. Sci. & Eng. 6:19(1962)
8. Freiser, H. & Biederman, Experiments on image quality in relation to the modulation transfer function and graininess of photographs, Photo. Sci. & Eng. 7:28(1963)
9. Meilgaard, M., Gail Civille, B. Carr, ***Sensory Evaluation Techniques, 2nd Edition***, CRC Press, Inc. Boca Raton, FL, 1991 ISBN 0-8493-4280-5

Biography

Peter Engeldrum, the president and founder of Imcotek, has over 30 years experience in color scanning, printing and imaging. Imcotek is a consulting firm that brings color and imaging science expertise to imaging companies. Peter has degrees in Imaging Science from the Rochester Institute of Technology (RIT), Rochester, NY, where he served as a faculty member in the Center for Imaging Science for six years.

Peter has published numerous technical papers on image quality and color imaging and has recently authored a book entitled ***Psychometric Scaling: A Toolkit for Imaging Systems Development***. Peter has taught courses on color and imaging subjects both publicly and under corporate sponsorship. He has served on the board of directors for the Munsell Color Science Laboratory at RIT and on several other corporate Technical Advisory Boards. An additional responsibility includes his role as Chief Scientist and co-founder of E-Color, a San Francisco company that provides a system for color correcting images for the Internet. Peter also holds several issued patents, and has pending patents, in imaging and display calibration.

Peter is presently the IS&T Visiting Lecturer and has been a member of IS&T since his undergraduate days at the Rochester Institute of Technology. Over his many years as an IS&T member, he has held offices in three Chapters and has been session chair at several conferences. In 1994 Peter was awarded the IS&T Service Award.