Teorie & Modelli, n.s., XII, 1-2, 2007 (199-210)

Why functional measurement is (still) better than conjoint measurement: Judgment of numerosity by children and adolescents

James Shanteau[*] (Manhattan), *L. Richard Pringle*[**] (Baltimore), & *Judy A. Andrews*[***] (Eugene)

Krantz, Luce, Suppes, and Tversky (1971, p. 445) used a reanalysis of the city-occupation study by Sidowski and Anderson (1967) to argue for the superiority of Conjoint Measurement (CM) over Functional Measurement (FM). Although Anderson (1982) refuted this argument and demonstrated the superiority of FM, Coxon (2006, p. 9) concluded, "the interaction (in Sidowski & Anderson) is an artifact of the assumption that the rating scale is interval level…(because) an order preserving additive representation is possible" (p. 7). Others (e.g., Dijkstra & Timmermans, 1997; Moskowitz & Itty, 2003; and Smith & Albaum, 2005) reiterated the superiority of CM based, in part, on the Krantz, et al. reanalysis of the city-occupation study.

In reply, Anderson argued that CM lacks power to reveal substantive interactions. Because CM uses a monotone (order preserving) transformation, systematic interactions may not show up. Another shortcoming of CM is the absence of an error theory – there is no way to know whether an ordinal violation is meaningful or not.

Advocates of CM cite new applications, especially in marketing (Dijkstra & Timmermans, 1997), health care (Sculpher et al., 2004), engineering (Furlan & Corradetti, 2006), accounting (Emery & Baron, 1979), and product development (Smith & Albaum, 2005). There are consultants (Orme, 2005) who advertise CM services to clients. As stated at the web site *www.answers.com* (19 March 2007) conjoint analysis "determines what combination of a limited number of attributes is most preferred by respondents…(to be used) in testing customer acceptance of product designs and assessing the appeal of advertisements". Further, "conjoint analysis re-

[*] Department of Psychology, Kansas State University.
[**] Department of Psychology, Goucher College.
[***] Oregon Research Institute.

search projects might represent a remarkable help for pharmaceutical companies during the development of new drugs, and if properly conducted they might even allow the estimation of product sale and market share". Finally, "an additive model allows the estimation of the total utility of different profiles or combinations of attributes, and consequently, it allows the identification of the optimal configuration of a new or existing product or service".

Software to conduct CM analyses is available from SPSS (http://www.spss.com/conjoint/) and Sawtooth Software (http://www.sawtoothsoftware.com). Thus, "CM is one of the success stories of Representational Measurement…thereby illustrating Coombs' dictum that more conservative measurement assumptions can nonetheless achieve a better-founded, justifiable and higher-level solution than pseudo quantification can" (Coxon, 2006).

The purpose of this paper is to demonstrate these claims about CM are misleading, i.e., CM provides an erroneous interpretation of a clearly visible pattern. These results support Anderson's contention that CM lacks power to reveal findings that are clear from a FM perspective.

Background

The manner in which children use inputs from multiple dimensions, including length, density, number, and size, to make numerosity judgments has received attention over the past decades (e.g., Brainerd, 1977; Cuneo, 1982; Gallistel & Gelman, 2000; Gelman, 1990; Piaget, 1968; Pufall & Shaw, 1972; Siegler & Booth, 2004). For instance, Cuneo (1982) showed that preschoolers combine the dimensions of length and density additively to assess numerical quantity. However, research has yet to show how children and adolescents combine the dimensions of size and density to form a judgment of numerosity. The purpose of this study was to apply FM and CM to analyze different models of size-density judgment across developmental stages.

Model analysis: Functional measurement analysis

There are two plausible models of the integration of size and density. An *adding model* of size and density is:

$$R = s + d + c \tag{1}$$

where R is the numerosity response and s and d are subjective values of size and density, respectively; c is a constant. When plotted in factorial fashion, Equation 1 predicts parallelism. Statistically, there should be main effects for both size and density, but no interaction.

A *multiplicative model* of size and density, the normatively correct model for this task, takes the form of:

$$R = s \cdot d + c\,. \tag{2}$$

When plotted, Equation 2 predicts a bilinear fan – a diverging series of straight lines. Statistically, size and density main effects should be significant; the size-by-density interaction should be significant and localized in the bilinear component (Shanteau, 1974).

Model analysis: Conjoint measurement analysis

According to CM logic, the first step is to reduce ratings of numerosity to rank orders. These rankings are subjected to two tests: *consistent ordering* and *cancellation*. The former is an ordinal test of parallelism, whereas the latter is an ordinal analysis for an interaction.

For the two model forms outlined above, CM predicts that Equation 1 will satisfy tests for consistent ordering and cancellation with no deviations. Equation 2 leads to the same predictions, since a multiplying model is monotonically equivalent to an adding model.

Measurement of subjective values

In addition to diagnosing integration rules, a second goal of this study is to use FM to measure the psychological values of size and density at each age. To facilitate this analysis, the physical spacing of size and density stimuli follows the pattern of 1 : 2 : 4 : 8 on both dimensions. By using common spacing, both dimensions contribute equally to variation of number. Logically, size and density scale values should have equivalent spacing for numerosity judgments. It is also noteworthy that the range of numbers (up to 256) is large enough to discourage direct counting.

Method

Ten first graders (mean age = 6.5 years), 10 fourth graders (mean age = 10.1), and 20 adolescents (mean age = 18.6) participated in the study.

The children were students at an elementary school in Manhattan, KS; both parents and school personnel gave permission for the study. The adolescents were students at Kansas State University who received course credit for participation. Each individual session lasted 20 to 30 min.

The stimuli consisted of 16 squares of dots from a 4 × 4 factorial design. The sizes of the squares were 12.5, 25, 50 and 100 cm^2. The densities of the dots in the squares were 0.32, 0.64, 1.28, and 2.56 dots/cm^2. Each array of dots was mounted on a white cardboard surface measuring 12.5 × 12.5 cm.

After seeing the dot array, the participant estimated number of dots using an unmarked 20-cm scale. Two extreme anchors (in view the entire time) defined the end-points of the scale. The low anchor had zero dots on a 1.25 × 1.25 cm array; the upper anchor had 400 dots on a 12.5 × 12.5 cm array. Subjects made estimates of the number of dots by using a sliding marker along an unmarked rating scale. The task for all subjects was judging whether the stimulus array had "as few as this (low anchor)" versus "as many as this (high anchor)". Because of the large numbers involved, subjects were told not to attempt to count the dots, but just to estimate.

As a manipulation check, participants judged the four stimuli along the positive diagonal of the 4 × 4 design. Because size and density increase jointly along the diagonal, judgments should increase monotonically regardless of the model used. All subjects passed this check of instructions and scale use. Each subject assessed the 16 stimulus arrays three times, with the stimulus cards shuffled before each replication.

Results

Graphical analyses

The left diagram in Figure 1 shows mean results for first graders. The four lines correspond to different density levels; the spacing on the horizontal axis represents subjective spacing of size values. The upper right point, therefore, gives the mean response (= 9.4) to 2.56 dots/cm^2 on a 100-cm^2 array. In comparison, the lower left point gives the mean (= 1.1) to 0.32 dots/cm^2 on a 12.5-cm^2 array.

Three findings are notable. First, the results are quite orderly, especially considering that average age of these participants was 6.5 years, i.e., numerosity judgment was meaningful for these children. Second, results plot as a series of more-or-less straight lines; both size and density appear to influence numerosity judgments. Third, the lines are close to parallel for the three lowest array sizes, with a divergence for the largest array size.

The central diagram in Figure 1 shows comparable results for fourth graders. Two observations are noteworthy. First, the pattern is similar to the first graders, i.e., orderly results with slightly diverging lines for the largest array size. Second, the spacing of the lines is wider, suggesting greater attention to the density factor.

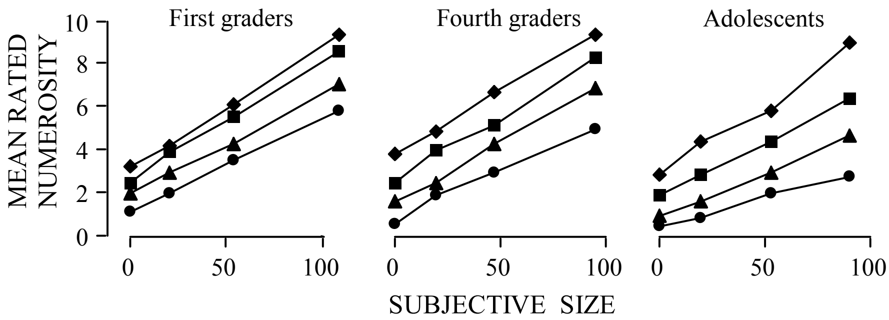

Figure 1. Mean rated numerosity of dots drawn on a square, plotted against subjective size of square for each of four dot densities [2.56 (◆), 1.28 (■), 0.64 (▲), and 0.32 (●)] for first and fourth graders and for adolescents.

The right diagram in Figure 1 plots the results for adolescents. In contrast to the children, these results are notably more divergent. This pattern is consistent with the normative multiplying model, which suggests the college students' numerosity judgments can be modeled by the product of size and density.

To confirm the graphical observations, the data were analyzed using both FM and CM techniques. The FM results appear first.

Functional measurement analyses

Because group analyses can easily mask individual-subject trends, each participant was analyzed separately. The results for all but one of the first graders revealed significant main effects for both size and density; for that subject, only density was significant. In analyzing the individual patterns of results, six were additive (i.e., no interaction), two were multiplicative (i.e., bilinear trend), and one could not be classified. Thus, the data patterns for most first graders were consistent with an adding strategy.

The results for all fourth graders yielded significant main effects for size and density. Only two had significant interactions, and, of these, one was bilinear. Thus like first graders, most fourth graders could be described by an adding strategy. All adolescents had significant main effects for size and density. Moreover, 15 also had significant interactions, with 12 of those localized in the bilinear component; the other three had patterns that appeared bilinear, but were not significant. Five adolescents showed no interaction and thus were consistent with an adding strategy.

To explore the findings further, an overall group × size × density analysis of variance was performed on the combined results for children vs. adolescents. All tests of significance were conducted at the 0.05 level. The analysis revealed significant main effects for group [$F(1, 38) = 21$), size [$F(3, 114) = 717$] and density [$F(3, 114) = 267$]. The interactions of group × density [$F(3, 114) = 4.8$], group × size [$F(3, 114) = 11$], and size × density [$F(9, 342) = 17$] and group × size × density [$F(9, 342) = 17$] were significant. A separate analysis of variance on first vs. fourth graders revealed main effects for size [$F(3, 54) = 326$] and density [$F(3, 54) = 78$]. The interaction of size × density was significant [$F(9, 162) = 4.6$]. There were no main effects or interactions involving group (grade level).

Scaling analyses

The general support for adding and multiplying means that the marginal means of the factorial design provide estimates of subjective scale values. For density, the actual stimulus spacing followed the ratio 1 : 2 : 4 : 8. For first graders the estimated spacing was 1.0 : 2.0 : 3.6 : 6.2; the first two values were set arbitrarily to ease comparison between subjective and objective values. For fourth graders, the estimated spacing was 1.0 : 2.0 : 3.4 : 5.75. For adolescents, the estimated spacing was 1.0 : 2.0 : 3.7 : 5.4. All three groups show compression of the density scale values for the higher values.

For size, the stimulus spacing also followed the ratio 1 : 2 : 4 : 8. For first graders the estimated spacing was 1.0 : 2.0 : 3.3 : 3.9. For fourth graders, estimated spacing was 1.0 : 2.0 : 3.5 : 4.5. For adolescents, estimated spacing was 1.0 : 2.0 : 3.9 : 5.5. Again, all three groups show compression of the scale values for the higher values, although adolescents were closer to the actual spacing.

Since both factors used the same 1 : 2 : 4 : 8 spacing, it is possible to compare the subjective spacing for each group between the factors. For first and fourth graders, the spacing was closer to accurate for density than for size, although the difference is most notable for first graders. For adolescents, in contrast, the correspondence of scale values was quite close for the two dimensions. This suggests that first graders have a greater underestimation of the ratio values for size relative to density. A similar, but less pronounced, effect appeared for fourth graders.

Conjoint measurement analyses

The two conjoint measurement properties of consistent ordering and cancellation provide necessary, but not sufficient, tests for an additive representation. All three groups satisfied both properties, with no violations. That is, an additive representation was satisfied for all groups, suggesting uniform support for an additive model. It is noteworthy that there was no evidence to support a multiplicative model for any group.

Discussion

The primary result of the FM analyses was that children tend to follow an adding rule (Equation 1) for size-density integration. Adolescents, in contrast, follow a multiplying rule (Equation 2). The FM analyses of scale values revealed a developmental shift in the subjective values for size and density. Adolescents tend to have more-or-less equivalent spacing of values, whereas the children had more compacted spacing for density relative to size. In comparison, CM analyses support adding for all groups, with no evidence of multiplying. These results suggest that CM is uninformative about the trends readily apparent in data plots and in FM.

Functional measurement vs. conjoint measurement

As illustrated here, FM provided an empirical basis for testing between two models identified originally: adding and multiplying. Children were consistent with adding and adolescents were consistent with multiplying. These trends were supported by statistical tests. There was also a trend in the scale values, with children showing wider spacing for size than for density; adolescents showed similar spacing for both dimensions

Meanwhile, CM found that all groups were consistent with an additive representation, i.e., CM was blind to the trends apparent to FM (and to the eye). This finding supports Anderson's (1982) argument that CM analyses lack power to diagnose model differences. It is deceiving to conclude that all groups follow an additive representation when there are clearly different integration strategies.

This weakness in CM shows up in other ways. Emery and Baron (1979) used simulations to conclude that CM has a bias toward an additive representation, regardless of what model actually generated the data. Moreover, CM articles tend to claim that "measure is possible" without providing actual estimates. This is a fatal weakness for any "measurement" approach. As Cliff (1992) concluded, CM is the "revolution that never happened" (p. 186).

Adherents of CM now use the term “Conjoint Analysis” (CA), which is distinct from CM in two respects. First, axiomatic tests of ordinal properties are not used; instead, global metric fits from standard curve fitting techniques are used. Second, measurement of values does not follow CM logic; rather, traditional statistical procedures, such as OLS (overall least squares) provide parameter estimates (Louviere, 1988).

Indeed many of the application successes of CA come from traditional metric scaling tools (Moskowitz & Itty, 2003). Green (1987) credits multidimensional scaling, not ordinal measurement theory, with being the inspiration of his often-cited 1971 paper (with Rao) on “Conjoint measurement for quantifying judgmental data”. Green goes on to state that “model validation and measurement reliability are also important areas (yet) to be studied” (p. 12). These are precisely the limitations that handicapped CM’s ability to diagnose the present results.

Implications for numerosity judgments

The present integration approach holds that judged numerosity is a function of perceived size and perceived numerosity. Accordingly, the strategy here was to investigate jointly the integration process and the perception process. FM proved fruitful in revealing developmental trends for both processes.

Similar to Anderson & Cuneo (1978) and Cuneo (1982), this study showed that children use two dimensions in their estimates of numerosity and that they combine them additively. Although Cuneo (1982) found that young children could combine dimensions multiplicatively if they were very simple arrays, the stimuli used in this study were too complex to allow direct estimation, e.g., counting. It appears that multiplicative processing of dimensions, such as density and size, does not emerge until later.

Conclusions

The present study both supports and extends previous FM research on developmental trends. To begin, there is now much evidence to support the additive-to-multiplicative shift in tasks where multiplying is normative; see similar findings in Anderson (1991) for payoff × probability judgments, estimations of area and volume, judgments of time, judgments of area, and adjustments on a balance scale. Therefore, in many domains where multiplication is appropriate, children initially follow an adding strategy before shifting to multiplying.

The present failure to find any evidence to support simpler strategies parallels prior research. Studies with children as young as 3 or 4 report no evidence of unidimensionality (Anderson & Cuneo, 1978). Although young children are consistent with an inappropriate adding rule, this means they have the capacity to integrate two (or more) dimensions.

In addition to a developmental trend for integration rules, there was a shift in the relative spacing of subjective values. For younger children, the spacing was more accurate for size than for density. There was no difference in relative spacing for adolescents. Such shifts suggest that developmental changes occur in scale values in addition to changes in combination rules.

In all, the insights from this study – developmental trends in both combination rules (from adding to multiplying) and scale values (from emphasis on size to coequal emphasis on size and density) – emerged only from FM analyses. In contrast, CM analyses were insensitive to these trends – from a CM perspective, there was no difference between these three groups. Thus, FM reveals what CM conceals.

References

Anderson, N. H. (1982). *Methods of information integration theory*. NY: Academic Press.

Anderson, N. H. (Ed.). (1991). *Contributions to information integration theory. Vol III: Developmental*. Hillsdale, NJ: Erlbaum.

Anderson, N. H., & Cuneo, D. O. (1978). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General, 107*, 335-378.

Brainerd, C. J. (1977). Effect of spatial cues on children's cardinal number judgments. *Developmental Psychology, 13*, 425-430.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science, 2,* 186-190.

Coxon, A. P. (2006). *Conjoint measurement*. ESRC, November 10 2006, City University, Edinburgh, Scotland.

Cuneo, D. (1982). Children's judgments of numerical quantity: A new view of early quantification. *Cognitive Psychology*, *14*, 13-44.

Dijkstra, J., & Timmermans, H. J. P. (1997). Exploring the possibilities of conjoint measurement as a decision-making tool for virtual wayfinding environments. In Y. T. Liu (Ed.), *Proceedings of computer aided architectural design in Asia* (pp. 61-72). Taipeh: Hu's Publishers.

Emery, D. R., & Barron, H. F. (1979). Axiomatic and numerical conjoint measurement: An evaluation of diagnostic efficacy. *Psychometrika, 44*, 195-210.

Furlan, R., & Corradetti, R. (2006). *Aspects of experimental design in the allocation-based conjoint analysis model. Proceedings of ENBIS*. Wroclaw, Poland: ENBIS.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Science*, *4*, 59-65.
Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and animate-inanimate distinction as examples. *Cognitive Science*, *14*, 79-106.
Green, P. E. (1987). Comments on "Conjoint measurement for quantifying judgmental data". *Current Contents, 9,* March 2, 1987.
Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, *8*, 355-363.
Gustaffson, A., Herrmann, A., & Huber, F. (2001). Conjoint analysis as an instrument of market research practice. In A. Gustaffson, A. Herrmann, & F. Huber (Eds.), *Conjoint measurement: Methods and applications* (pp. 5-46). Berlin: Springer.
Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement. Vol 1*. New York: Academic Press.
Louviere, J. J. (1988). *Analyzing decision making: Metric conjoint analysis.* Beverly Hills, CA: Sage Publications.
Moskowitz, H. & Itty, B. (2003). Jumpstarting product development: Competitive analysis and conjoint measurement in the cosmetic industry. *Journal of Advertising Research, 43*, 62-77.
Orme, B. (2005). *Getting started with conjoint analysis*. Madison, WI: Research Publishers.
Piaget, J. (1968). Quantification, conservation, and nativism. *Science, 162*, 976-979.
Pufall, P. B., & Shaw, R. E. (1972). Precocious thought on number: The long and the short of it. *Developmental Psychology, 7*, 62-69.
Sculpher, M., Bryan, S., Fry, P., DeWinter, P., Payne, H., & Emberton, M. (2004). Patients' preferences for the management of non-metastatic prostate cancer: Discrete choice experiment. *British Medical Journal, 328*, 382-385,
Shanteau, J. (1974). Component processes in risky decision making. *Journal of Experimental Psychology, 103*, 680-691.
Sidowski, J. B., & Anderson, N. H. (1967). Judgments of city-occupation combinations. *Psychonomic Science*, *7*, 279-280.
Smith, S. M., & Albaum, G. S. (2005). *Fundamentals of Marketing Research*. Thousand Oaks, CA: Sage Publications.

Abstract

Supporters of Conjoint Measurement (CM) continue to cite an argument from Krantz, Luce, Suppes, and Tversky (1971) from a reanalysis of the city-occupation study by Sidowski and Anderson (1967). Anderson (1982) refuted this argument and demonstrated the superiority of Functional Measurement (FM); one of the arguments made by Anderson is that CM lacks power to reveal substantive interactions. Still, Coxon (2006, p. 9) recently concluded, "the interaction (in Sidowski &

Anderson) is an artifact of the assumption that the rating scale is interval level…(because) an order preserving [monotone] additive representation is possible" (p. 7). In the present study, first grade children, fourth grade children, and adolescents assessed judgments of numerosity based on size (area) and density. FM analyses revealed that adolescents followed the normative multiplicative rule, while children were generally additive. CM analyses, in contrast, were insensitive to the different integration rules followed by children and adolescents, i.e., an additive representation existed for all ages. Thus, an important developmental trend that is apparent both graphically and statistically in FM analyses disappeared in CM analyses. This not only supports Anderson's arguments about the superiority of FM techniques, it provides a relevant counter-example to the CM arguments based on the Sidowski and Anderson study.

Riassunto

I sostenitori del Conjoint Measurement (CM) continuano a citare un argomento di Krantz, Luce, Suppes, and Tversky (1971) riguardante una rianalisi dello studio città-occupazione di Sidowski and Anderson (1967). Anderson (1982) rifiutò tale argomento e dimostrò la superiorità della Misurazione Funzionale (MF); uno degli argomenti di Anderson è che il CM manca di potere per rivelare interazioni reali. Inoltre, Coxon (2006, p. 9) ha recentemente concluso che "l'interazione (in Sidowski & Anderson) è un artefatto della assunzione che la scala di valutazione è a livello di intervallo…(perché) una rappresentazione additiva [monotona] che conservi l'ordine è possibile" (p. 7). Nel presente studio, bambini di prima e di quarta elementare e adolescenti hanno giudicato la numerosità basata sulla grandezza (area) e la densità. Analisi basate sulla MF hanno rivelato che gli adolescenti seguono la regola normativa moltiplicativa, mentre in generale i bambini sono additivi. In contrasto, le analisi basate sul CM sono risultate insensibili alle differenti regole di integrazione seguite dai bambini e dagli adolescenti, cioè esisteva una rappresentazione additiva per tutte le età. Perciò, una tendenza di sviluppo importante che è evidente sia graficamente che statisticamente nella analisi basata sulla MF scompare nella analisi basata sul CM. Ciò non solo supporta gli argomenti di Anderson circa la superiorità delle tecniche di MF ma fornisce anche un controesempio rilevante agli argomenti del CM basati sullo studio di Sidowski and Anderson.

Acknowledgment. This research was supported in part by a National Institute of Child Health and Human Development grant HD11857 to the first author and a National Institute of Mental Health Training Grant award MH08359 to the third author. Preparation of the manuscript was supported in part by an Air Force Office of Scientific Research grant 06NL156 to the first author. We wish to thank Norman Anderson, Ruth Phelps, and Michael Troutman for helpful comments on an earlier version of this manuscript.

Addresses. James Shanteau, Department of Psychology, Bluemont Hall 492, Kansas State University, Manhattan, KS 66506-5302 USA (shanteau@ksu.edu). L. Richard Pringle, Department of Psychology, Goucher College, 1021 Dulaney Valley Road, Baltimore, MD 21204, USA (rick.pringle@goucher.edu). Judy A. Andrews, Oregon Research Institute, 1715 Franklin Boulevard, Eugene, OR 97403, USA (judy@ori.org).