Episodic Organization in Discourse and Valid Measurement in the Sciences*

John W. Oller Jr1 and Liang Chen2
1University of Louisiana at Lafayette; 2University of Georgia at Athens

ABSTRACT

The argument here concerns the episodic organization of discourse and measurement in the sciences. It is developed in two parts. First, empirical effects of episodic organization in discourse processing are demonstrated. The impact on first and second language users is shown by demonstrating loss in comprehension when episodic organization is disrupted. Empirical evidence is also summarized showing that the depth of comprehension of episodic organization is sensitive to relatively minute elements of surface form such as the names of persons and places in the discourse. It can be inferred that the ecological validity of language tests and measures of discourse processing in general is enhanced by respecting episodic organization and reduced by disrupting it. The second part of the argument concerns the deeper question of why episodic organization enhances the ecological validity of language tests and discourse processing measurements in general. Looking into this question from the point of view of Peircean logic, we are led to a surprisingly general inference: that valid measurements of any kind depend on the episodic organization of ordinary experience.

WHAT IS EPISODIC ORGANIZATION?

The phrase episodic organization calls to mind the sequence of a series of episodes in a story-line as in a film or a novel. From the empirical point of view episodic organization is commonly manifested in the chronological arrangement of narratives and in ordinary reports of event sequences. In the sciences, experiments and reports of experiments are, as we will show, utterly dependent on the episodic organization of experience in general.

*Address correspondence to: John W. Oller Jr, Hawthorne Regent’s Professor, Department of Communicative Disorders, University of Louisiana at Lafayette, PO Box 43170, Lafayette, LA 70504-3170, USA. E-mail: joller@louisiana.edu

0929-6174/07/1402-30127S16.00 © 2007 Taylor & Francis
Episodic organization is often referred to with reference to schemata, scripts, frames, and the like (e.g. Anderson & Pearson, 1984; Carrell, 1987, 1988; Chen & Graves, 1995, and their references).

In its most common manifestation, episodic organization arises from the fact that when a number of distinct events happen to occur in a certain order, say, 1, 2, 3, etc., they are also likely to be recalled and reported in that order. The idea is an old theme but a current topic of discussion. Spinoza (1674, p. 451) overstated the case in saying that “the order and connection of ideas is the same as the order and connection of things” but the theme has been reiterated by Graesser, Millis, & Zwaan (1997, p. 173): “...the order of mention should match the order of occurrence of events”. Ohtsuka & Brewer (1992, p. 323) said “elements of language [i.e. the series of constructions used in discourse] are...isomorphic with the universe [i.e. the series of events referred to by those constructions]”. Enkvist (1981, p. 98) called the agreement of these interlinked sequences “experiential iconicism” and it is true that in a reported series of events arranged in a strict chronology, there is an iconicity (a strict analogy) between the sequence of events (E) and the sequence of constructions (C) reporting those events.

In case the constructions, C_1, C_2, etc., appear in the same order as the events, E_1, E_2, etc., we find the simplest manifestation of the referring or pragmatic mapping relation, π, where the Cs are arranged in the same order as the Es, i.e. C_1πE_1, C_2πE_2, etc., and where every event is reported at least once and no event is reported more than once. However, this simplest π mapping sequence is certainly not the only one that is possible. As commonly noted, the Es of any narrative can be reported in any order, though arrangements deviating from the original chronology will ordinarily require special marking in the discourse (Pike, 1954).

Nonetheless, the idealized case, which is evidently the one that Spinoza (1674) and the others mentioned had in mind, is worth examining further. If all else is held equal and if each event is reported once and only once, as the string of triads (CπE_1, CπE_2, etc.) grows in length the number of possible sequences of n CπE triads, in any given narrative can be expressed as n! Thus, for two events, two orders of reporting are possible in the idealized form; for length three, six orders are possible; for four, 24 and so on, though only one of these sequences will conform to the principle of isomorphism. Nevertheless, as Spinoza and others have argued, there is a powerful preference for the episodic organization where the chronology of the constructions is isomorphic with the chronology of the reported events.
For example, Ohtsuka & Brewer (1992) found that texts violating the canonical chronological order of narratives, e.g. texts with flash forward events or flashbacks, were more difficult to process. In fact, those with flash forwards were extremely disruptive and off-putting. Comprehension of flash forwards was reduced to near chance levels. Similarly, Alderson et al. (2000) found that when they asked EFL students to put a jumbled text consisting of six sentences following a given lead-in back into its original order, subjects barely performed above chance levels.

Why is this? At first blush we may simply say that it is because of the fact that the chronological order is simplest. This is a good answer, but it will turn out not to be complete. A more complete answer, as shown in the second part of this paper, will show that the simplicity of the chronological order is a logically necessary one without which even nominal distinctions between objects could not be achieved. Nor could any kind of ordinal measurement be achieved without the presupposition of the episodic organization of experience.

To begin with, it is useful to note that dual chronologies of an event sequence and a sequence of constructions (e.g. clauses in discourse) can only reduce to a single chronology in the case where they are indeed, in keeping with Spinoza and others, isomorphic. It is unsurprising, therefore, that the tendency toward isomorphism should prevail in narratives. As we have argued elsewhere (Chen & Oller, 2003), the tendency to prefer the event chronology in narratives appears not only to be universal across languages and cultures (Lin, 1989) but also can be seen in the preferred subject verb object (SVO) order in simple declarative transitive sentences. Not only is the SVO order the most common across languages (Greenberg, 1987), but it reflects an underlying chronology of distinct events that may be analysed as follows:

1. S refers to, i.e. is \( \pi \) mapped onto, the agent that typically conceives of and initiates the action to be reported.
2. The V is \( \pi \) mapped to the action of that the agent subsequently carries out.
3. The O is \( \pi \) mapped to some other referent that is impacted by the action referred to in the verb.

Typically, a narrative-like chronology of events can be discerned in the SVO sequence: the agent (1) initiates the act before it (2) takes place and
only after that act is underway can it (3) impact its object. Hence, the SVO order, in simple transitive constructions, reflects an underlying chronology of events. While any order of reporting, even at the sentence level, is theoretically possible, that is, SOV, VOS, VSO, OSV, OVS are all not only possible but attested (Dryer, 1997 and his references), it is clear from empirical studies that orders conforming best to known or inferred chronologies of actual event sequences are more likely to be used, easier to produce, and easier to recall.

PART 1. EPISODIC ORGANIZATION ENHANCES VALIDITY IN LANGUAGE TESTS AND DISCOURSE MEASURES

Effects of various aspects of episodic organization have been shown in language tests and measures of discourse processing (Zwaan, 1996; Munte et al., 1998; Oller et al., 2005). Chen & Oller (2003; also Badon et al., 2005) have observed that more episodically organized materials produce higher reliability and validity and more interpretable measures. Better structured materials produce more discrimination across distinct proficiency levels, yield higher task motivation, and higher indexes of validity. Moreover, disrupting episodic organization has the opposite effects.

In what follows, we extend the analysis of certain data from Chihara et al. (1994). They used a counterbalanced repeated measures design with 201 Japanese adults studying English as a foreign language at the Osaka YMCA (71 beginners, 66 intermediate, and 64 advanced students) plus 41 native speakers of English in an introduction to linguistics course at the University of New Mexico. For the sake of the trend analysis reported here, the distinct proficiency levels were reduced by random deletion of subjects to equal cell sizes of 40 in each cohort (per Maxwell & Delaney, 1990, pp. 221–235) or 160 cases in all.

Subjects
For the analyses that follow there were 40 college students at each of four levels – beginning EFL, intermediate EFL, advanced EFL, and native speakers of English. All 160 subjects were adults above 18 years of age of whom 120 were EFL students at the Osaka YMCA in Japan and 40 were undergraduate students in introductory linguistics at the University of New Mexico in Albuquerque.
Method
Two similar narratives of approximately 350 words each were used. Test A was about a young man, Joe, preparing to go off to college, and Text B was about another young man, Nicholas, setting out to visit relatives in Greece. Both texts were presented as written cloze tasks by deleting roughly every 6th word in one of two conditions, either in the original chronological arrangement (sequential) or with their sentences of the story rearranged in a shuffled sequence (scrambled; cf. full texts in Oller & Jonz, 1994, pp. 145–146). Each cohort of 40 subjects were assigned randomly to four blocks with 10 subjects in each block defined by the order of presentation of both condition and text in each instance: (i) Text A sequential, Text B scrambled; (ii) B scrambled, A sequential; (iii) A scrambled, B sequential; and (iv) B sequential, A scrambled. Items were the same across conditions; only the order of sentences was changed. The design involved two completely crossed between-subjects variables (the four proficiency levels and the four blocks) and three within-subjects variables each nested within the other two: order (first or second) nested within text (A and B) which in turn was nested in the main variable of interest, condition (sequential versus scrambled).

Results
To assure that the method was not contaminated by differences between blocks or an order effect, an omnibus analysis of variance was applied to the randomly created blocks. Averaging over condition, text, and order, expressed as percentages times 100, the means for each block respectively were 36.92, 37.82, 37.94, 38.63 as contrasted against the grand mean of 37.83. The contrasts between these means were null ($F_{3, 156} = 0.055, p = 0.90$) and there was no proficiency by blocks interaction.

As predicted, contrasts between sequential and scrambled texts increased with increasing proficiency in English as predicted and as shown in Tables 1 to 3 and Figure 1 below. Table 1 reports the mean percentage correct as scored by the exact-word method, i.e., subjects had to guess the very word deleted from each segment of text. The mean difference of 6.19 between the mean on the sequential texts (40.92) versus the mean on the scrambled texts was significant ($F_{1, 159} = 47.79, p < 0.0001$) as was the contrast across proficiency levels ($F_{3, 156} = 867.33, p < 0.0001$) and the interaction between condition by proficiency level ($F_{7, 152} = 17.47, p < 0.0001$).

Table 2 reports the breakdown of the differences across proficiency levels showing the contrasts of interest. Following Maxwell & Delaney...
Table 1. Mean percentages of exact-word scores between groups by condition.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Sequential</th>
<th>Mean</th>
<th>s</th>
<th>Scrambled</th>
<th>Mean</th>
<th>s</th>
<th>Difference</th>
<th>Mean</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning EFL</td>
<td>40</td>
<td>19.03</td>
<td>7.04</td>
<td></td>
<td>17.90</td>
<td>8.20</td>
<td></td>
<td>1.12</td>
<td>10.16</td>
<td></td>
</tr>
<tr>
<td>Intermediate EFL</td>
<td>40</td>
<td>32.13</td>
<td>6.88</td>
<td></td>
<td>27.77</td>
<td>7.60</td>
<td></td>
<td>4.36</td>
<td>10.45</td>
<td></td>
</tr>
<tr>
<td>Advanced EFL</td>
<td>40</td>
<td>40.94</td>
<td>9.41</td>
<td></td>
<td>32.75</td>
<td>8.06</td>
<td></td>
<td>8.18</td>
<td>13.18</td>
<td></td>
</tr>
<tr>
<td>Native English</td>
<td>40</td>
<td>71.59</td>
<td>8.86</td>
<td></td>
<td>60.51</td>
<td>9.71</td>
<td></td>
<td>11.08</td>
<td>9.01</td>
<td></td>
</tr>
<tr>
<td>Grand means</td>
<td>160</td>
<td>40.92</td>
<td>21.05</td>
<td></td>
<td>34.73</td>
<td>17.97</td>
<td></td>
<td>6.19</td>
<td>11.36</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Sequential versus scrambled condition by proficiency level.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Mean difference</th>
<th>s</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Beginning EFL</td>
<td>1.12</td>
<td>10.16</td>
<td>1, 39</td>
<td>0.50</td>
<td>0.5</td>
</tr>
<tr>
<td>2. Intermediate EFL</td>
<td>4.36</td>
<td>10.45</td>
<td>1, 39</td>
<td>7.14</td>
<td>0.025</td>
</tr>
<tr>
<td>3. Advanced EFL</td>
<td>8.18</td>
<td>13.18</td>
<td>1, 39</td>
<td>15.82</td>
<td>0</td>
</tr>
<tr>
<td>4. Native English</td>
<td>11.08</td>
<td>9.01</td>
<td>1, 39</td>
<td>62.04</td>
<td>0</td>
</tr>
<tr>
<td>Beginners versus intermediates</td>
<td>3.24</td>
<td>10.30</td>
<td>1, 78</td>
<td>3.85</td>
<td>0.1</td>
</tr>
<tr>
<td>Intermediates versus advanced</td>
<td>3.82</td>
<td>11.74</td>
<td>1, 78</td>
<td>4.03</td>
<td>0.05</td>
</tr>
<tr>
<td>Advanced versus natives</td>
<td>2.90</td>
<td>10.90</td>
<td>1, 78</td>
<td>2.57</td>
<td>0.25</td>
</tr>
<tr>
<td>Beginners versus advanced</td>
<td>7.06</td>
<td>9.64</td>
<td>1, 78</td>
<td>19.87</td>
<td>0</td>
</tr>
<tr>
<td>Intermediates versus natives</td>
<td>6.72</td>
<td>11.00</td>
<td>1, 78</td>
<td>13.81</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3. Trend analysis for differences due to episodic organization by proficiency level.

<table>
<thead>
<tr>
<th>Group</th>
<th>Beginning EFL</th>
<th>Inferred EFL</th>
<th>Advanced EFL</th>
<th>Native English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean difference</td>
<td>1.123</td>
<td>4.36</td>
<td>8.185</td>
<td>11.08</td>
</tr>
<tr>
<td>Contrast coefficients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Quadratic</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Cubic</td>
<td>-1</td>
<td>3</td>
<td>-3</td>
<td>1</td>
</tr>
<tr>
<td>ANOVA source</td>
<td>SS</td>
<td>df</td>
<td>MS</td>
<td>F</td>
</tr>
<tr>
<td>Between groups</td>
<td>9106</td>
<td>3</td>
<td>3035</td>
<td>23.69</td>
</tr>
<tr>
<td>Linear</td>
<td>9083</td>
<td>1</td>
<td>9083</td>
<td>70.88</td>
</tr>
<tr>
<td>Deviation from linearity</td>
<td>23</td>
<td>2</td>
<td>12</td>
<td>0.09</td>
</tr>
<tr>
<td>Quadratic</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>0.04</td>
</tr>
<tr>
<td>Cubic</td>
<td>18</td>
<td>1</td>
<td>18</td>
<td>0.14</td>
</tr>
<tr>
<td>Within subjects</td>
<td>156</td>
<td>128</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(1990, p. 461) the $F$-ratios pertaining to contrasts at individual proficiency levels are based on the formula:

$$F = \frac{n\bar{D}^2}{s_D^2}$$

where $\bar{D}$ is the mean observed difference between sequential and scrambled texts. For the contrasts between levels, the $F$-ratios consist of the difference between the pair of $\bar{D}$s in question in each instance divided by the pooled variance within each the two proficiency levels.

Figure 1 plots the trend line for difference scores as reported in Tables 1 and 2. As individuals increase in proficiency, as predicted they benefit more from episodic organization. Beginners are hardly, if at all, able to take advantage of episodic organization while intermediates and higher proficiency levels become increasingly able to do so. While the contrast between adjacent levels of proficiency (e.g. beginners versus
intermediates, intermediates versus advanced, and advanced versus natives) fail in two of three instances to achieve significance at \( p < 0.05 \), the contrast between beginners and advanced EFL students, and between intermediates and native speakers of English is significant in both cases even with the Bonferroni correction for multiple comparisons (i.e. where the preset \( \alpha \) is divided by the number of contrasts examined per Maxwell & Delaney, 1990, p. 177, or in this instance 5). As predicted, the effect of episodic organization increases with each succeeding level of proficiency and, as shown in Table 3 below, the increasing degree of contrast owed to episodic organization reveals a significant linear trend (\( F_{\text{linear}} 1, 156 = 70.88 \)). The omnibus test for deviation from linearity was not significant and there were no significant quadratic or cubic components.

Discussion

The foregoing results show that second language learners and native speakers use episodic organization to fill in missing elements in a cloze task. They are increasingly able to make use of such information as they gain in proficiency. Other research shows that the depth of comprehension is sensitive to relatively minute elements of surface form such as the familiarity of the proper names in the text. Chihara et al. (1989) hypothesized that relatively small changes in the surface forms of texts would make them more comprehensible to Japanese speakers learning English as a foreign language (EFL). Changing names to more familiar forms improved overall scores by about 6% on the texts studied. Al-Fallay (1994) amplified the method and extended the findings to additional texts with Arabic learners of EFL. The effects of episodic organization generalized to English proficiency tests that subjects had never seen before. Sasasaki (2000) followed up to see just which items in the text about Joe going off to college were affected by the surface form changes made by Chihara et al. (1989). She changed (1) Klein’s (a department store in New York) to Daiei (a well-known department store in Japan); (2) Joe (the name of the central character) to Hiroshi; (3) Bellevue (a town in the story) to Kyoto; (4) Ned to Kazuo; (5) Plaintown to Nara; (6) Lizzie to Haru; (7) state to prefecture; and (8) kissing to hugging. As predicted the mean scores on the original text and the modified version differed significantly (by 9.57% for two groups of Japanese learners of EFL, \( F 1, 29 = 10.10, p < 0.005 \)). Chi-square analyses (\( df = 1, 60 \)) showed contrasts at \( p < 0.05 \) in all the proper nouns except Joe/Hiroshi which occurred in nearly all the sentences of the text. Sasasaki tested subjects
individually and invited them to reflect on their answers. The original
text reads as follows, with the numbers in parentheses showing the cloze
items (given as blanks of a standard length in the test situation) and the
deleted words (shown in upper case letters here) that the test-taker
should guess:

...at his (12) MOTHER’S suggestion, he had his father’s tailor (13)
MEASURE him for a suit. He didn’t (14) HAVE a suit made, though,
because his (15) FATHER wouldn’t let him order one. “You’re (16)
STILL growing, son,” he said. “You’re growing (17) SO fast that you’d
outgrow a suit (18) IN no time. Buy yourself a pair (19) OF slacks and a
sport jacket. Klein’s (20) HAS such a large selection that I’m (21)
SURE you will find something you like (22) THERE.” Joe’s father
always suggested Klein’s for (23) CLOTHES Joe went to Klein’s in
order (24) TO please his father but he didn’t (25) FIND anything that
he liked there so (26) HE went to another store to buy (27) THE slacks.
He took them out of (28) THE box as soon as he got (29) HOME so
that his father wouldn’t notice (30) WHERE they came from . . . .

In the elicited protocol, referring to item (20), the student reported:

Student: And he bought a sports jacket and other things.
Sasaki: Uh, huh.
Student: Hmm . . .
Sasaki: And then?
Student: . . . I don’t know whether this Klein is a man or a woman,
but . . .
Sasaki: Uh, huh. Well, OK, what happened about that person, Klein?
Student: Well, Klein, Klein, Klein . . .
Sasaki: Yes?
Student: His father . . . I don’t know what his father asked Klein,
but . . .

When Klein’s is taken as the name of a person rather than a
department store, the comprehension of the entire event sequence is
compromised.

In Sasaki’s study, cloze items that contributed to significant contrasts
between the original and modified versions of the text all consisted of
proper nouns. Each of these reduces to a triad consisting of a surface
form \((F)\) pragmatically mapped \((p)\) onto a person or place \((P)\). If any such a triad, \(FpP\), is incorrectly construed, it is easy to see how the misconstrual will interfere with subsequent constructions in which that triad is a component. In retrospect it seems obvious that changing the surface forms of referring terms to more familiar ones would make it easier for learners to interpret them, and this is exactly what Sasaki (2000) found. However, it is not obvious that enabling correct construal of referring terms will logically give an ecologically more valid measure of the discourse processing at issue. Yet, this follows from the fact that measures grounded in valid reference must be more valid than those grounded in misconstruals of referential relations. If the key arguments of any construction cannot be pinned down referentially, higher syntactic, semantic, and pragmatic relations also become less determinate. Therefore, improving the comprehension of the referring relations ought to make it possible to get more valid information from discourse. Sasaki’s results support these more subtle inferences.

Her original text resulted in lower reliability than the modified (more comprehensible) version. Contrasts in the respective reliabilities were significant both by Guttman’s split-half method of estimating reliability, 0.63 versus 0.70 \((F_1, 29 = 6.09, p = 0.0197)\) and by the more robust Cronbach-\(\alpha\) method, 0.76 versus 0.89 \((F_1, 29 = 23.20, p < 0.0001)\). It follows that if episodic organization enhances meaningful discrimination across subjects it must tend to improve the validity as well. The research reviewed elsewhere (see Oller et al., 2005) supports this generalization very broadly with respect to discourse based measures of comprehensibility, production, comprehension, recall, and/or the generalizability (i.e. learnability) of linguistic materials whether spoken, heard, read, written, or otherwise applied in verbal reasoning.

**PART 2. WHY DOES EPISODIC ORGANIZATION ENHANCE VALIDITY?**

Supposing for the sake of argument that the foregoing evidences have been correctly understood, why does episodic organization enhance the validity of discourse measures? Obviously, there ought to be (and evidently there is) a tendency for the order of constructions in discourse to conform to the orderly arrangements of things in experience. As we have already shown above, that sort of mapping arrangement is the simplest and requires the
least effort (along the lines of Zipf, 1949; also Pike, 1954). However, this explanation presupposes that events in experience are arranged chronologically without saying why they are arranged that way. More particularly, why should episodic arrangements of any kind be important to validity in measurements of any kind? In studying these questions from the point of view of a general theory of signs (Oller et al., 2005) or from a general logic of relations (Peirce, 1867, 1897), some surprising results emerge. For one, we discover that no valid nomination (naming) of any identity whatever could be achieved apart from the episodic organization of experience. Thus, episodic organization must be presupposed because it is logically indispensable. It is not merely the simplest of all possible arrangements of signs with respect to a sequence of events, but (surprisingly) it is a necessary logical prerequisite to the discovery of any valid sign relations whatsoever. For another surprising result, the logical analysis to follow, which is grounded in a general theory of signs, shows that every ordinal relation (every transitive relation) utterly depends on the sort of iconic transitivity manifested in the isomorphism of a sequence of events reported in an identical sequence of propositions. In the end it will come out that episodic organization is an absolute foundational prerequisite to any kind of valid measurement whatsoever.

Stevens (1968, p. 850) defined four distinct classes of scales, nominal, ordinal, interval, and ratio, on the basis of “the group of transformations” under which each “scale remains invariant”. Countless other types are possible, as he noted, but it can be shown logically that naming or numbering (referring as with a nominal distinction or scale) and setting the names or numbers in a logical order (as in an ordinal scale) are both foundational requirements for the building up of all higher types scales or measures applied in the sciences. He noted, for instance, explained as follows: “A nominal scale admits of any one-to-one substitution . . . [and an] ordinal scale can be transformed by any increasing monotonic function” (Stevens, 1968, p. 850). As examples he proposed the numbers assigned to players on a team and “the hardness scale determined by the ability of one mineral to scratch another” (Stevens, 1968, p. 850). We will show that higher (more complex) scales than nominal and ordinal ones, e.g. interval and ratio scales, must meet the foundational requirements for nominal and ordinal scales at a minimum. Therefore, by showing that these foundational scales (measures) depend utterly on the episodic organization of experience will assure us that higher scales of more complex kinds share that dependency.
We discover from the vantage point of a general theory of signs (Peirce, 1867, 1897; Oller, 1993, 1996) that a demonstrative nominal judgment, e.g. saying that a certain object is of a given kind, is peculiar in differentiating its object from all others that might have been singled out for attention. A classic example of such a demonstrative nominal judgment is the kind seen in a valid application of any proper name or referring term in discourse, e.g. the kind that turned out to be important to performance in Sasaki’s study above. From a logical point of view, such a nominal demonstrative dynamically divides the universe of possible objects of attention into two classes, the one named (or numbered) and all similars that might have been named or numbered but were not. It can be seen that such a distinction quantitatively is isomorphic with the distinction between any pair of distinct numbers, e.g. say, between 1 and 0, or between any pair of distinct names, or descriptive signs. Thus, the distinction itself that is contained in a demonstrative nominal judgment hardly qualifies as a scalar judgment of any kind. A nominal “scale” constitutes the crudest sort of order, but it can be shown that any ordinal judgment whatsoever is utterly dependent on the cruder sort of discriminative judgment contained in a nominal judgment.

The two elements together, reference by naming (or numbering) together with the ordinal relation contained in the distinction between whatever is referred to as distinct from whatever is not referred to, however, logically provide necessary elements of all possible measurement scales that have any hope of validity. This can be seen immediately from the fact that a name (or number) or any ordinal value applied “willy nilly” (indiscriminately) to any and all possible objects or to no objects at all cannot be a valid measure of anything whatever. To achieve the slightest modicum of validity or truth, as argued coherently by Borsboom, Mellenbergh, & van Heerden (2004), also Badon et al. (2005), the measure must at the very least single out to some extent the object of measurement.

From the vantage point of sign theory (Oller, 2005; Oller et al., 2005), any valid nominal judgment (an act of reference) involves the sort of triadic relation described as \( S \pi O \) where \( S \) is a sign (a name or number arbitrarily applied) through \( \pi \) which is an active mapping relation (by an intelligent action of some measurer, counter, or device) and \( O \) is the object nominated (numbered or measured). In this triadic relation we find all three of the major systems of signs Peirce (1867, 1897) described in his
general logic of relations – icons, indices, and symbols. To start with there are at least three icons:

1. \( S \), must be noticeable and must have an iconic form (a surface form) displayed in some manner or other, e.g. as a written or uttered name or number distinct from other icons.
2. \( O \), the object of measurement, must also be known through some iconic manifestation that is noticeable in some manner or other as by perception.
3. \( \pi \), the act of measuring itself or pointing out of the \( O \) must be a noticeable act manifested perceptually as a distinct icon including both the icon of the \( O \) and the icon of the \( S \). This mapping relation can only be displayed through some act of pointing, looking, or otherwise indexing the association of the \( S \) with its \( O \).

It can also be demonstrated logically that all three of the foregoing elements are necessary to any valid nominal measurement. Without an actual \( O \), the \( S \) could only \( \pi \) map a fiction (an invented or imaginary datum) and would fail to that extent to be a valid measurement. Without the \( S \), \( O \) would remain less than a raw datum. It could be no more than an unmeasured (unnamed, unknown) element in the continuum of countless possible data that might be singled out for attention. Finally, without \( \pi \), some specific act of association connecting \( S \) with \( O \), there would be no way to arbitrarily nominate \( O \) as distinct from the multitude of other objects that might conceivably be nominated with by the designation \( S \). This last claim is certain on account of the fact that any conventional sign, \( S \), because it is only associated with its \( O \) by convention, cannot be applied to \( O \) at all except through some act establishing or instantiating the conventional application (Oller, 2005).

Next, to the very extent that any conventional application of any \( S \) is arbitrary, a nominal scale must be valid if applied according to its convention and not valid if it violates that convention. The name (or number) either is or is not applied according to its conventional use. But how shall this be determined? Clearly, the standard of validity in nominal scales is nothing but the established naming, referring, or representational convention. For instance, either subject number 14 (by the name of Joe Smith, say) is correctly identified or not. But how is the arbitrary naming or numbering of any given object to be determined and known? How can the convention be established in the first place?
As soon as we address this problem, which may be termed the nominal validity problem, we discover a necessary connection between validity and episodic organization. For any \( S \) to be a valid mark (name or number) assigned to any \( O \), it is necessary within the limits of the conventional application for the same \( O \) to receive the same \( S \) each time. For the arbitrarily assigned \( S \) to serve its purpose of identifying the \( O \) of interest, it is necessary that the \( S \) be applied consistently to that \( O \) and not to others. But how is \( O \) to be recognized on distinct occasions as the same one to which \( S \) was previously applied, and, how will similar \( O \)s be distinguished from the one identified previously as \( S \)? Clearly, arbitrary assignment of a distinct \( S \) to every appearance of any \( O \) would merely add complexity indiscriminately and would serve no purpose of valid measurement (naming or numbering) whatever.

Keeping in mind the nominal validity problem, with ordinal scales something more is required. In the case of an ordinal scale the marks on the scale must not only be assigned so that objects are identified more or less correctly but are also ordered. All ordinal scales have the property of logical transitivity where if \( A \) bears a certain relation to \( B \) and \( B \) bears that same relation to \( C \), then \( A \) must also bear that relation to \( C \). It is clear that the ordinal validity problem involves nominal validity plus something more. The ordinal validity problem is to arrange the measured objects according to an ordered relation of signs. Clearly, the ordinal problem can only be solved to the extent that the prior nominal problem has been solved. It turns out that neither problem can be solved at all unless ordinary experience is episodically organized.

With the two described validity problems in mind, suppose the sign-user (measurement expert) were to assign distinct \( S \)s (names or numbers) based on sensory impressions of \( O \)s according to Ockham’s razor and Leibniz’s principle of the identity of indiscernibles. Combining their maxims, the Ockham-Leibniz rule is as follows: if objects seem different, they should be named differently, otherwise, the same. But it can be shown that these principles cannot work at all without reference to the dynamic temporal order of perceptual events in ordinary experience. The case can be made by considering the universal limits on the representational power of icons (as noted clearly first by Peirce). If isolated and unrelated perceptual icons had to be judged one-by-one (like timeless snapshots, immediate sensations, unsequenced isolated events or isolated constructions representing unconnected events or states), the nominal problem could not be solved at all. The insurmountable barrier to its
solution is the fact that all icons resemble each other and also resemble all other objects (and all other icons). As a result, it would never be possible by comparing icons to determine with any reasonable certainty whether any given icon represented similar but different objects, or the same object in different states or at different times.

But, suppose we introduce the ordinary episodic organization of experience where perceptual icons are associated either with relatively constant identities distributed over space and time within coherent boundaries in any given field of observation, or with distinct bounded entities that maintain their distinct boundaries as they are distributed across time and space. When bounded entities move as bodily units, they take their boundaries with them as they simultaneously separate themselves from distinct other bounded objects in time and space. In such an episodically organized context, the nominal validity problem can be solved by the application of the Ockham-Leibniz rule. Moving objects retain their boundaries as they move. Similarly, objects distributed in space relative to each other retain their distinct boundaries over time as can be discovered by movements of the perceiver. By tracking bounded objects as they move over space and time, the problem of object identity can be partially solved. At least the objects present at the same time and perceivable in a particular place can be distinguished by this method. However, to solve the problem of nominal validity for objects that come and go from the field of perception requires more than merely associating dynamic indexes with those objects. Clearly, a dynamic index, which must be attached at one end to an observer, cannot pursue a moving object beyond the boundaries of the field of perception of that observer. So, how is the object to be known as the same or different if it re-enters the field of perception of the same observer but at a later time?

Suppose the object is assigned an arbitrary conventional label on some perceptual encounter. If this assignment can be independently checked on subsequent encounters – say, by asking for a photo ID, a fingerprint, or some other identifier that can be independently checked – the nominal validity problem can, just to that extent, be solved. If this can be done on the basis of the ordinary episodic organization of experience, then, it is clear that the ordinal validity problem can be addressed in similar fashion. The surprising conclusions stated earlier follow by irrefragable logic: all more complex measures, and, therefore, all valid measures of any kind must meet the requirements of nominal validity to a high degree as well as the requirements of ordinal validity.
It follows that all measures whatsoever depend on the episodic organization of experience.

One further step in the argument is also needed to show fully its generality. As Peirce (1867, 1897) demonstrated in his general logic of relations, all valid inferences depend on valid transitive relations. What is more, any valid ordinal relation, not only chronological ones, can provide a suitable basis for valid logical inferences and therefore for the connectedness of discourse. Although the most common transitive relations commonly manifested in discourse are (1) chronologies, we also find (2) cause-effect chains, (3) sequences of actions in the pursuit of a goal, and (4) a vast variety of asymmetrical transitive relations such as being greater than (or less than), to the left (or right) of, above (or below), etc., along with the doubly transitive relation of equality, where if \( A = B \), and \( B = C \), \( C = A \), etc. Any rule or relation whatever that provides a basis for inference (whether necessary, probable, or merely possible) must also depend on episodic organization.

For all the foregoing reasons, if all else is held equal, discourse processing tasks of all sorts that conform better to the expected episodic organization, ought not only to be more comprehensible, easier to recall, easier to learn from, and in all ways easier to process, but also more authentic, more genuine, and more ecologically valid when used as measures of any sort.

**CONCLUSION**

Since the ordinary episodic organization of experience is thus essential to the attainment of validity in nominal and ordinal scales, and since all higher scales of measurement actually applied to objects, events, or states of affairs in space and time depend ultimately on valid nominal and ordinal judgments, it must be expected that discourse materials, as well as scientific measures of all kinds, that respect the episodic organization of experience will yield higher validities in general. Logically, a general theory of signs thus provides hope for the unified validity sought by Messick (1993, 1998) and alluded to, in principle, by Stevens (1968), also in generalizability theory (Cronbach et al., 1963; Shavelson et al., 1989). Evidently, both reliability and validity are grounded in a kind of ecological generalizability (validity) that is essentially dependent on the episodic organization of ordinary experience.
ACKNOWLEDGEMENTS

This paper was presented at the Quantitative Linguistics Conference 2003 (QUALICO) in Athens, Georgia at the University of Georgia, Continuing Education Conference Center on May 28, 2003. The authors are grateful to all of the participants who commented on the presentation and especially to Patrick Juola and Harald Baayen. Any errors that remain and the ideas expressed, however, are our own.

REFERENCES


