The S–N Scale & the SP Question

Wayne D. Mitchell
Palmer, AK

Data from a large best-fit study show that although SPs tend to give more item responses opposite to their best-fit preferences than other types do, NJs are far more frequently misclassified on the S–N scale than SPs are.

ABSTRACT

Data from a national sample compared MBTI scored results to validated type for 2,521 cases. SPs gave more opposite item responses on the S–N scale compared to other temperaments or S–N/J–P letter combinations. However, NF and NT, and NJ and NP letter groups were the ones most frequently misclassified on the S–N dichotomy. SPs were most frequently misclassified on the J–P and E–I scales, whereas SJs were most frequently misclassified on the T–F scale and the E–I scale. Administrator bias did not appear to have influenced SP best-fit types.

INTRODUCTION

Several reports over the past few years have concluded that the MBTI S–N scale elicits an inordinate number of intuitive responses from SPs (Berens, 1997; Berens & Robb, 1997; Rogers, 1997). Most of these reports have been based on anecdotal rather than empirical data. Although anecdotal evidence may be suggestive of interesting research questions, it is inadequate for drawing scientific conclusions. Advances in both the hard sciences and the behavioral sciences depend on the replicability of observations and measurements by different researchers. Replication requires standardized procedures for the collection and analyses of data. Without such standardized procedures, the developing knowledge in a field is subject to the kinds of observer bias that have been documented for years in all sciences (Hempel, 1952).

This is not to denigrate the human observer or anecdotal observations. In appropriate contexts, such as counseling individual clients, these may have as much or more value as the best

"objective" measurements. However, the human observer and anecdotal observation are inadequate tools when one seeks to generalize one's observations and conclusions to a widely used instrument, to certain classes of respondents, or to other practitioners in the field. Human beings may unintentionally influence people and events. They may also perceive and remember events selectively. Anecdotal, nonsystematic, and nonquantified observations are neither replicable nor generalizable to other settings, samples, or observers. It is for this reason that scientific research favors the use of systematic observations and analyses that can be replicated by independent investigators.

THE SP QUESTION

A recent line of more systematic, quantitative inquiry has indicated that presenter and administration context may influence the accuracy of scored MBTI results (Bathurst, 1999). Bathurst found that intuitive administrators tended to elicit a greater number of changes from S to N than vice versa when MBTI scored results were validated in the feedback process with respondents. His findings suggest that even best-fit SPs may unknowingly be misclassified on the S–N scale if feedback has been given by an intuitive administrator. For these reasons, it seemed appropriate to study "the SP question" in a more systematic and rigorous way than has been done to date. To this end, data from a nationwide best-fit study were analyzed for the purpose of answering several key questions concerning the response patterns of people with the SP preferences.

QUESTIONS ADDRESSED IN THIS STUDY

1. Is there a greater tendency for best-fit SPs to give more opposite responses to the MBTI's S–N items compared to other best-fit temperament groups or comparable best-fit two-letter preference combinations? An "opposite response" refers to a response to an item that is opposite to a person's best-fit preference on the scale to which the item belongs.

2. If SPs tend to give more opposite responses compared to other groups, to what extent are their scores on the S–N scale affected?

3. Compared to their scored types, do people who verify their best-fit preferences as SP tend to be misclassified on the S–N scale to a greater degree than other types?

4. Is there any evidence in this study that SPs might have been influenced to validate as Ns rather than as Ss?

METHOD

Sample. To obtain data for a study having different purposes than the present one, Consulting Psychologists Press (CPP) asked MBTI practitioners to contribute data on clients for whom best-fit as well as scored type could be obtained. Practitioners contributing data to the study were recruited through mailed announcements to large corporate customers, calls to experienced people who kept all of their type data, and referrals from CPP's customer service representatives. Practitioners were offered a discount on Form J answer sheets as an incentive to submit their data. Roughly 36 practitioners submitted data to the study. Best-fit type was determined by the practitioners' usual and preferred practices. Respondents rated their confidence in their best-fit type on a scale from 1 to 5, with 1 representing "not at all confident" and 5 representing "very confident." Confidence

ratings were made for the four preferences separately as well as for whole type. Best-fit type was defined as "This is the type that you think fits you best after hearing the interpretation. This type may be the same as your MBTI reported type, or may be a different type that you think fits you better" (Hammer, personal communication, March 5, 2000). A total of 2,521 cases were available for analysis.

Analyses. The data were analyzed using SPSS 9.0. Counts of opposite responses, ANOVAs, and other statistics were all run using self-assessed best-fit type as the basis for identifying respondent whole type and all four-letter preferences. The number of opposite responses on the S–N scale and continuous scores on that scale were the dependent variables. Temperament groups and the four S–N/J–P letter combinations were the independent variables.

RESULTS

Type frequencies of the best-fit types are shown in Table 1, which compares the scored results to the best-fit results. The selection ratios in this table reflect the degree to which scored types and preferences occurred more or less frequently than the corresponding best-fit types and preferences. It is clear from the SRTT table that for no type or letter combination was the scored type either significantly more frequent or less frequent than the best-fit type. However, these results yield only a very general picture. The questions on which this study focused require more detailed analyses.

Question 1. ***Is there a greater tendency for best-fit SPs to give opposite responses to S–N items compared to other best-fit temperament groups or to comparable best-fit two-letter preference combinations?*** Table 2 shows the mean number of S and N responses for the four temperament groups and the NP and NJ letter combinations. Several observations from the table bear on this question. A one-way ANOVA with number of opposite responses as the dependent variable and temperament as the main factor was significant ($p < .001$), so a multiple comparison test (MCT) was run to identify the differing groups. The mean number of item responses in the N direction was significantly higher ($p < .001$) for the SP group (11.73) than for the SJ group (9.85). The mean number of item responses in the S direction was virtually the same for the NF (7.17) and the NT (7.02) groups. The SP and SJ groups both differed significantly from the NF and NT groups ($p < .001$).

The second observation from Table 2 is that in round numbers, the NF and NT groups on average gave opposite responses on two to four fewer items compared to the SP and SJ groups. Furthermore, the SP group on average chose only 3.5 more S than N responses. The SJ group on average chose 7.1 more S than N responses. Both of these were considerably lower than for the NF and NT groups, which on average chose 12 more N than S responses.

The lower part of Table 2 reveals a difference between intuitive types that is not apparent in looking only at temperament groups. On average, the NJ group gave nearly two more S responses than the NP group did. An ANOVA testing differences between the SJ, SP, NJ, and NP groups was significant ($p < .001$). The post hoc MCTs revealed that all four of these letter combinations differed significantly from each other ($p < .001$). Thus, when temperament groups were compared, the sensing temperaments were significantly different from the intuitive temperament groups in their mean number of opposite responses, and they were also

Table 1. **Type Distribution of Scored Types in Best-Fit Study Sample and SRTT Comparison With Best-Fit Types.**

N = 2,521 + = 1% of *N* *I* = Selection Ratio Index *$p<.05$ **$p<.01$ ***$p<.001$

The Sixteen Complete Types

ISTJ *n* = 267 (10.6%) *I* = 1.03 + + + + + + + + + + +	ISFJ *n* = 257 (10.2%) *I* = 0.96 + + + + + + + + + +	INFJ *n* = 127 (5.0%) *I* = 1.05 + + + + +	INTJ *n* = 94 (3.7%) *I* = 1.3 + + + +
ISTP *n* = 71 (2.8%) *I* = 0.97 + + +	ISFP *n* =123 (4.9%) *I* = 1.04 + + + + +	INFP *n* = 231 (9.2%) *I* = 0.94 + + + + + + + + +	INTP *n* = 76 (3.0%) *I* = 1.10 + + +
ESTP *n* = 72 (2.9%) *I* = 0.95 + + +	ESFP *n* =156 (6.2%) *I* = 0.89 + + + + + +	ENFP *n* = 266 (10.6%) *I* = 0.92 + + + + + + + + + + +	ENTP *n* = 86 (3.4%) *I* = 0.91 + + +
ESTJ *n* = 196 (7.8%) *I* = 1.11 + + + + + + + +	ESFJ *n* = 265 (10.5%) *I* = 1.07 + + + + + + + + + + +	ENFJ *n* = 153 (6.1%) *I* = 1.12 + + + + + +	ENTJ *n* = 81 (3.2%) *I* = 1.03 + + +

Jungian Types (E)

	n	%	*Index*
E–TJ	256	10.2	1.08
E–FJ	384	15.2	1.09
ES–P	252	10.0	0.90
EN–P	382	15.2	0.92

Jungian Types (I)

	n	%	*Index*
I–TP	142	5.6	1.04
I–FP	365	14.5	0.97
IS–J	528	20.9	0.99
IN–J	212	8.4	1.04

Dominant Types

	n	%	*Index*
Dt. T	398	15.8	1.07
Dt. T	749	29.7	1.03
Dt. T	780	30.9	0.96
Dt. T	594	23.6	0.96

Dichotomous Preferences

E	1275	(50.6%)	*I*=1.00
I	1246	(49.4%)	*I*=1.00
S	1407	(55.8%)	*I*=1.01
N	1114	(44.2%)	*I*=0.99
T	943	(37.4%)	*I*=1.03
F	1578	(62.6%)	*I*=0.99
J	1440	(57.1%)	*I*=1.04
P	1081	(42.9%)	*I*=0.95

Pairs and Temperaments

IJ	745	(29.6%)	*I*=1.01
IP	501	(19.9%)	*I*=0.99
EP	580	(23.0%)	*I*=0.91
EJ	695	(27.6%)	*I*=1.09
ST	606	(24.0%)	*I*=1.03
SF	801	(31.8%)	*I*=0.99
NF	777	(30.8%)	*I*=0.98
NT	337	(13.4%)	*I*=1.01
SJ	985	(39.1%)	*I*=1.03
SP	422	(16.7%)	*I*=0.95
NP	659	(26.1%	*I*=0.94
NJ	455	(18.1%)	*I*=1.07
TJ	638	(25.3%)	*I*=1.05
TP	305	(12.1%)	*I*=0.98
FP	776	(30.8%)	*I*=0.94
FJ	802	(31.8%	*I*=1.04
IN	528	(20.9%)	*I*=1.00
EN	586	(23.2%)	*I*=0.98
IS	718	(28.5%)	*I*=1.00
ES	689	(27.3%)	*I*=1.02
ET	435	(17.3%)	*I*=1.02
EF	840	(33.3%)	*I*=0.99
IF	738	(29.3%)	*I*=0.98
IT	508	(20.2%)	*I*=1.03

Wayne D. Mitchell, The S–N Scale and the SP Question

significantly different from each other. However, when the four S–N/J–P letter combinations were compared, they were all significantly different from each other in the mean number of opposite responses. An analysis by temperament groups thus masked significant differences between the intuitive groups that were compared here.

Question 2. ***If SPs tended to give more opposite (N) responses compared to other groups, to what extent were their scores on the S–N scale affected?*** The far right column of Table 2 reveals an 8.4-point difference in the average continuous scores of the SP and SJ temperament groups, with the SP group being closer to the S–N scale midpoint. This difference was statistically significant ($p < .001$) in a one-way ANOVA. There was no difference between the mean continuous scores of the intuition temperament groups. If the analysis is limited to temperament groups, SP continuous scores are the only ones skewed toward the midpoint of the S–N scale.

However, the lower two rows in Table 2 reveal a 7.5-point difference between the mean continuous scores of the NP and NJ groups, with the NJ mean score being closer to the midpoint of the S–N scale. This difference was also significant in a one-way ANOVA ($p < .001$). The pattern for continuous scores thus was similar to that for the number of opposite responses to S–N items. That is, when temperament groups alone were compared, the SJ and SP groups differed significantly from each other and from the NF and NT groups, but the latter two did not differ. When the four S–N/J–P letter combinations were compared, all four groups differed significantly from each other.

When numbers of opposite responses are considered, SPs had the highest mean among all other groups. When continuous scores are considered, the mean for NJs was closer to the midpoint than for any other group listed in Table 2. The greater number of opposite responses for SPs did not result in their S–N continuous scores

Table 2. **Mean Number of S and N Responses to S–N Items and Mean Continuous S–N Scores for Various Letter Combinations.**

Best-Fit Letter Combinations	Number of Cases	Mean Number of Item Responses in S Direction	Mean Number of Item Responses in N Direction	Mean S–N Continuous Score
Temperament Groups				
NF	792	7.17	19.69	120.91
NT	333	7.02	19.90	120.99
SP	443	15.19	11.73	82.63
SJ	956	17.02	9.85	74.22
N/J–P Letter Combinations				
NP	698	6.46	20.40	123.78

being closest to the midpoint. Despite their choosing opposite responses to half as many items as SPs did, the continuous score for NJs was closer to the midpoint than for any other group listed in Table 2. These differences in continuous scores reflect the well-known correlation between S and J, and between N and P.

To see whether the aggregated analyses by temperament and S–N/J–P groups might be obscuring important whole type differences (Mitchell, 2000), the four whole types within each temperament group were compared against each other in an ANOVA. No significant differences were found between the types within any of the temperament groups on either number of opposite responses or continuous scores. Similarly, no significant differences were found between the types within the NJ, NP, SJ, or SP groups.

Question 3. ***Compared to their scored types, do people who verify their best-fit preferences as SP tend to be misclassified on the S–N scale to a greater degree than other types?*** Table 3 shows the most frequent and second most frequent incorrectly scored scales for the four temperament groups and the four S–N/J–P letter combination groups. Among the temperament groups, NF and NT were most often incorrectly scored as S. SJs were most often incorrectly scored on the T–F scale, whereas nearly a fourth of the SPs were incorrectly scored as Js. Among the four S–N/J–P letter combinations, NJs and NPs were most often incorrectly scored as Ss.

Table 3. **Most Frequent Incorrectly Scored Letter Preferences, Best-Fit Temperament, and S–N/J–P Letter Pair Combinations.**

	Most Frequent Incorrectly Scored Letter Preferences*		Second Most Frequent Incorrectly Scored Letter Preferences*	
Temperament Groups				
NF	S–N	16.7%	J–P	12.4%
NT	S–N	18.6%	T–F	17.1%
SJ	T–F	14.6%	E–I	11.4%
SP	J–P	23.9%	E–I	16.7%
S–N/J–P Letter Pairs				
SJ	T–F	14.6%	E–I	11.4%
SP	J–P	23.9%	E–I	16.7%
NJ	S–N	25.5%	J–P	21.2%
NP	S–N	12.2%	T–F	11.7%

**Note*: Percentages of each best-fit group that were mis-scored on the indicated scale are listed. Thus, of all best-fit SPs, 23.9% were mis-scored on the J–P scale; of all best-fit NJs, 25.5% were mis-scored on the S–N scale. The percentages in these columns were calculated on the total number of the corresponding best-fit types. For example, of all best-fit NFs, 16.7% were misclassified on the S–N scale (i.e., were scored as S, but validated their true preference as N).

Fully one-fourth of best-fit NJs were incorrectly scored as Ss, and over a fifth of them were incorrectly scored as Ps.

The difference between NJs and NPs in Table 3 is notable. Twice as many NJs (25%) incorrectly scored as Ss compared to NPs (12%). Almost the same percentage of NJs were also misclassified by the J–P scale (21%). Further analysis revealed that only 3.5% of best-fit NJs were misclassified on both S–N and J–P, so the overlap between these two groups of misclassified NJs is fairly small.

From these data, it is clear that the intuition temperament groups and the intuition S–N/J–P letter pair groups are most frequently misclassified as S. The SP and SJ groups are most frequently misclassified on the J–P, T–F, and E–I scales. Consequently, the answer to this question is no: SPs are not misclassified on the S–N scale to a greater degree than other types.

Question 4. *Is there any evidence in this study that SPs might have been influenced to validate as Ns rather than Ss?* There was no way to test for this directly, as the types of administrators were not available. However, it is possible to make some reasonable inferences based on the existing data. Bathurst (1999) found that administrators preferring intuition tended to elicit more changes towards intuition than expected by chance. If we assume that the majority of administrators providing data for this sample preferred intuition, then based on Bathurst's results, we would expect to see more changes from S to N than from N to S when

Table 4. **Percentage of Best-Fit Respondents Whose Preference for S or N Differed From Their Scored Results.**

	Percent Changed From Scored N to Best-Fit S*	Percent Changed From Scored S to Best-Fit N*
Temperament Groups		
SJ	5.7%	
SP	13.1%	
NF		16.7%
NT		18.6%
S–N/J–P Letter Combinations		
SJ	5.6%	
SP	13.1%	
NP		12.2%
NJ		25.5%

**Note:* The percentages in these columns were calculated on the total number of the corresponding best-fit types. Thus, of all best-fit SJs in the sample, 5.7% scored as N but validated as S.

comparing best-fit to scored preference on the S–N scale. If we compare the SJ and SP groups in the top half of Table 4, this *appears* to be the case.

However, the bottom half of Table 4 shows that this conclusion is unjustified. If respondents who scored as SP had been influenced to choose N as their best-fit type, they would appear in the NP row of the table under the column labeled "% changed from scored S to best-fit N." The percentages of changers shown in the SP and NP rows are virtually equal. This tells us that the percentage of respondents changing from scored SP to best-fit NP is the same as the percentage changing from scored NP to best-fit SP. Even if we assume that all administrators in this sample preferred intuition, these results indicate that any influence they may have had on self-validation of type was not biased against SPs.

Table 5 reports these same percentages for all Ss and all Ns and grouped by pairs of types differing only on the S-N dichotomy. The contrast between the percentages aggregated over all Ss and Ns compared to pairs of Ss and Ns shows the importance of examining data for individual types. The aggregated data give one impression, but the paired type data show a quite different picture.

We see, for example, that 15.9% of best-fit INTPs scored as S but validated as N, compared to 19.2% of ISTPs who scored as N but validated as S. If administrator bias had pushed this sample to validate more often as N rather than S, we would expect the percentage of best-fit INTPs changing from scored S to validated N to be larger than the percentage of best-fit ISTPs changing from scored N to validated S, for example. Similarly, for the other pairs in the table, we would expect that the percentage of best-fit intuitive types changing from S to N would be larger than the percentage of best-fit sensing types changing from N to S. This was true for all of the SJ and NJ pairs, but only for the ESFP/ENFP pair. For the other three SP/NP pairs, a larger percentage of SPs changed from scored N to best-fit S than NPs changed from scored S to best-fit N. So the answer to this question is no: It does not appear that administrator bias unduly influenced SPs to validate as N. If administrator bias was present in these data, it primarily affected the NJ types.

DISCUSSION

The best-fit data in this study showed that anecdotally reported misclassification of SPs as Ns on the MBTI instrument (Berens, 1997; Berens & Robb, 1997; Rogers, 1997) was not confirmed when tested rigorously with numerical data on a large sample. Using best-fit data, the groups most frequently misclassified on the S–N scale were the intuitive temperament groups NF and NT, and the NJ and NP letter pair groups. The types of administrators in this sample were not known. Bathurst's (1999) recent work on administrator bias, and his observation that 86% of his administrators were intuitive (personal communication, April 7, 2000) raised a concern as to whether the results presented here were artifacts of that bias. Yet even under assumptions that all administrators were intuitive and that they influenced respondents to validate as N rather than S, the data clearly showed that on the S–N scale, the percentage of misclassified SPs was no greater than the percentage of misclassified NPs. Based on these data, there is no justification for claiming that the MBTI instrument misclassifies SPs as intuitives or that administrator bias induced "true" SPs to validate as Ns in this study.

Best-fit SPs tended to give more opposite responses on the S–N scale compared to the other temperament groups and the other S–N/J–P letter

Table 5. **Percentage of Best-Fit Respondents Whose Preference for S or N Differed From Their Scored Results.**

	Percent Who Scored N and Changed to S*	Percent Who Scored S and Changed to N*
Best-Fit Letter Preference		
S (n = 1,399)	8.0	
N (n = 1,125)		17.2
Best-Fit Type		
ISTP	19.2%	
INTP		15.9%
ISFP	13.6%	
INFP		10.1%
ESTP	15.8%	
ENTP		11.7%
ESFP	9.1%	
ENFP		13.2%
ISTJ	7.3%	
INTJ		22.0%
ISFJ	4.5%	
INFJ		28.9%
ESTJ	6.8%	
ENTJ		25.3%
ESFJ	4.4%	
ENFJ		25.0%

**Note*: The percentages in these columns were calculated on the total number of the corresponding best-fit types. Thus, of all best-fit ISTPs, for example, 19.2% were mis-scored on the S–N scale (i.e., were scored as N, but validated their true preference as S).

pair combinations. This result confirms anecdotal impressions reported elsewhere (Berens, 1997; Berens & Robb, 1997; Rogers, 1997). However, the mean continuous scores of best-fit NJs were closer to the midpoint than any other temperament group or S–N/J–P letter pair. The scale on which SPs were most frequently misclassified was the J–P scale; the second most frequent misclassification for SPs was on the E–I scale. The intuitive temperament groups (NF and

NT) as well as the NJ and NP letter pair groups were most frequently misclassified on the S–N scale. Thus, the trends observed in patterns of opposite responses did not produce higher rates of misclassification of SPs on the S–N scale. If there is a problem with that scale, it appears to be with how it classifies intuitive types, especially NJs, not with how it classifies SPs.

Without knowing administrator type, it was impossible in the present study to explore further the possible sources of classification error for NJs on the S–N scale. The fact that there was little overlap between NJs who were mis-scored on the S–N and J–P scales suggests that there may be two sources of error and that they may operate independently of each other. Administrator bias is certainly a source of error that deserves further study. At this stage of our understanding, it is not clear whether any such bias might operate in favor of certain preferences (e.g., towards N and P) or away from certain preferences (e.g., away from S and J). What is clear from the present study is that if it exists, such bias does not adversely affect the classification of SPs.

Different practitioners prefer different styles and methods of presenting type and validating best-fit type for their clients. Validation of type is not always a replicable or infallible process (Rytting, 1998). The present study had no control over this process. Had data been provided by only a small number of practitioners, this might have been a major weakness, because best-fit type would have reflected only a few limited and perhaps very similar methods of type validation. However, data were provided from 36 different practitioners working in diverse settings ranging from small academic settings to therapist caseloads and corporate training programs. This diversity effectually randomized feedback methods and precluded any single technique or orientation from unduly distorting the best-fit validation results for respondents in this study.

Two other points are worth noting. The first is that the results presented here indicate that it is misleading to say that "a vast number of SPs select N rather than S responses on the MBTI" (Berens, 1997, p. 43). With no data or systematic analysis presented to test that observation, the implication is that large numbers of SPs are therefore mistyped as Ns. The data from this study show that although SPs do tend to give more opposite responses to S–N items than other S–N/J–P letter combinations, they are clearly not misclassified on the S–N scale to the degree that intuitive types are. As one of the practitioners concerned with the SP question noted, ". . . perception is an active process and the frameworks that we hold are like colored lenses we look through that color our perception of events" (Berens, 1998, p. 37). This is precisely why conclusions subject to the rigors of scientific methodology are more trustworthy than casual observations. The data collected and analyzed in this study were not anecdotal, but systematically collected, quantitative, and therefore replicable—crucial standards for valid scientific knowledge.

Personal biases affect not only our observations and the interpretations we make but may actually impact the events we are trying to study or observe. Administrator bias did not appear to affect SPs in any significant way in this study. However, if the unconscious bias of administrator type affects the validation of best-fit type (Bathurst, 1999), one wonders how sizeable might be the bias effect caused by a strong and consciously held belief in the existence of an SP misclassification problem.

A second point emerges from the finding that NJs are much more frequently misclassified on the S–N scale than are SPs. This result was

completely obscured in analyses by temperament groups. It was only apparent in looking at the data by the S–N/J–P letter pairs. The latter groups are consistent with the MBTI model and emerge directly from type theory as a logically consistent, mutually exclusive grouping based on a subset of type preferences. The four temperament groupings, on the other hand, are not consonant with type theory (Rytting, 1996). Conceptually they subsume different things: "Each of the four whole temperament patterns has essential elements that provide information not in the type pattern—a map of a different kind and at a different level" (Nardi & Berens, 1998, p. 9).

Not only does the conceptual content of the models differ in scope, but the models also divide that content differently. Two of the temperament groups (SP and SJ) mix together Ts and Fs and differ only by the J–P dichotomy. The other two temperament groups (NF and NT) mix together Js and Ps and differ only by the T–F dichotomy. Temperament groups can be meaningfully contrasted when considered from the perspective of temperament theory. They cannot be meaningfully contrasted from the perspective of type theory. The reason is that when translated into type letters, temperament groups are identified on the basis of different pairs of dichotomies. From the perspective of type theory, how meaningful is it to contrast two groups, one of which contains both Fs and Ts, and the other of which contains only Ts? Or how meaningful is it to compare a group containing only Ps with a group containing both Js and Ps?

The MBTI instrument was developed as a means of identifying Jung's types. It was not developed with the intent of identifying temperaments. For this reason, efforts to overlay the temperament model onto the raw structure of the four MBTI scales create confusion on a number of theoretical and empirical levels. As Berens (1996) noted, "there is a close relationship between the theories and instruments used to access those theories. They cannot be separated" (p. 9). The NJ findings in this study are a striking illustration of this relationship between theories and instruments and of the hazards that arise in ignoring that relationship. It appears that concerns over the presumed misclassification of SPs by the Indicator derive from an ill-advised effort to use the Indicator to identify temperament personality styles.

It is true that there is some overlap in the content of temperament and the MBTI types; but "overlap" does not mean "the same as," a point that has been too often overlooked. The combined S and P preferences on the MBTI instrument do not identify "artisans" and were never intended to do so by the Indicator's author. Whatever overlap exists between temperament theory's artisan type and the Indicator's combined S and P preferences is a coincidence. The overlap emerged when two creative and insightful people, Myers and Keirsey, were trying to describe limited aspects of the same phenomenon, viz., human personality.

Some practitioners prefer using multiple models in helping clients reach a better understanding of their personalities (Nardi, 1999). This approach may be helpful so long as the models and their differences are clearly distinguished. But it does a major disservice to clients, to type and temperament models, and to instruments, when the method developed to measure one model (MBTI instrument) is presented as if some parts of it were isomorphic to the other model (temperament).

It also impedes further development of our understandings of type when temperament's lack of fit with MBTI results is then interpreted as a

problem with the Indicator. The confusion is compounded when anecdotal results are cited as evidence of a problem with the S–N scale, but no hard, scientific data are offered to support the claims being made. Science has exponentially increased human knowledge using methods of thought, data collection, and analysis that help ensure that our personal biases minimally distort what we learn through research. Some recent philosophical perspectives advocate that research should be located within an individual's personally configured universe and that scientific research is only one of several ways of knowing that needs to be connected to other ways (Bentz & Shapiro, 1998). This perspective emphasizes experiential, connected awareness over the more detached, systematic, and replicable procedures of most sciences. Although there is ample evidence that science is a socially influenced activity (Berger & Luckmann, 1967; Kuhn, 1970), this influence does not negate the value of rigorous, systematic, empirical research. The results of this study have called into serious question the notion that the S–N scale of the MBTI instrument has excessively misclassified SPs. Findings from the present study point to areas in which sound research offers us a chance to expand, rather than obscure, what we know of type.

REFERENCES

Bathurst, J. (1999). *Best fit vs. reported type: Demographic and environmental effects on the accuracy of the MBTI.* Paper presented at the 13th biennial international meeting of the Association for Psychological Type, Phoenix, AZ.

Bentz, V. M., & Shapiro, J. J. (1998). *Mindful inquiry in social research.* Thousand Oaks, CA: Sage Publications.

Berens, L. (1996). Type and temperament. *Bulletin of Psychological Type, 19,* 8–9.

Berens, L. V. (1997). Extraverted sensing, the MBTI, and the artisan terperament. *Bulletin of Psychological Type, 20,* 42–43.

Berens, L. V. (1998). What is true type, anyway? *Bulletin of Psychological Type, 21,* 37–38.

Berens, L. V., & Robb, J. (1997). How to tell intuiting from extraverted sensing (Se). *Bulletin of Psychological Type, 20,* 15–18.

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality.* Garden City, NY: Doubleday & Company, Inc.

Hempel, C. G. (1952). Fundamentals of concept formation in empirical science (*International encyclopedia of unified science,* Vol. 2.) Chicago: University of Chicago Press.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Mitchell, W. D. (2000). Cautions regarding aggregated data analysis in type research. *Journal of Psychological Type, 53,* 19–30.

Nardi, D. (1999). Using multiple models to explore type. *Bulletin of Psychological Type, 22,* 22–26.

Nardi, D., & Berens, L. V. (1998). Wizards in the wilderness: Searching for true type. *Bulletin of Psychological Type, 21,* 7–9.

Rogers, S. (1997). *Se and the MBTI: Findings of dissertation research.* Paper presented at the 12th biennial international meeting of the Association for Psychological Type, Boston, MA.

Rytting, M. (1996). Measuring type and temperament. *Bulletin of Psychological Type, 19,* 6–7.

Rytting, M. (1998). The validity of validated type. *Bulletin of Psychological Type, 21,* 4–6.

CONTACT

Dr. Wayne D. Mitchell
HC03 Box 8450E
Palmer, AK 99645-9406
(907) 745-1730
wmitch@matnet.com