Personality Testing in the Church of Scientology: Implications for Outcome Research

John H. Wolfe[*]
San Diego, California

Abstract. Many fields of modern society require scientific proof of effectiveness before new methods can be widely accepted, as in clinical trials for new drugs, educational evaluation for teaching approaches, and outcome studies for psychological interventions. Previous outcome studies on the results from Scientology services are reviewed and found to be inconclusive. The paper is devoted to the question of whether the existing data base of several thousand case histories could be used for outcome studies. The existing data contain personality test scores on the Oxford Capacity Analysis (OCA) administered before and after scientology services. A detailed analysis of the OCA demonstrates that it was derived from the Johnson Temperament Analysis (JTA), a psychological test of poorly documented validity, by paraphrasing its items, copying its scoring weights and transforming its test norms, with some alterations. It was concluded that the OCA is presently unsuitable for outcome studies, but that this situation could change if additional research could demonstrate that the OCA had validities comparable to other personality tests. For future use, it is recommended that an entirely new version of the OCA be constructed with completely original items, simplified scoring weights, and empirically derived norms, and that its validity and reliability be demonstrated prior to implementation.

Keywords: Scientology, outcomes, OCA, Oxford Capacity Analysis, validation

One of the most controversial of the new religions to arise in the last century, *Scientology* defines itself as "an applied religious philosophy dealing with the study of knowledge, which through the application of its technology, can bring about desirable changes in the conditions of life." (Hubbard, 1975, p. 370). A precursor, and now a subfield of Scientology, was *dianetics* (Hubbard 1950), which is a psychotherapy supposed to help an individual eliminate psychosomatic illnesses, mental aberrations and impediments to realizing his/her full potential, by removing or desensitizing mental image pictures of traumatic incidents. The application of dianetic or Scientology technology to an individual with the purpose of achieving improvements that s/he desires is called *auditing* or *processing*. In early years, L. Ron Hubbard, the founder of dianetics and Scientology, often claimed that they were sciences, based on scientific research, but later described them as philosophical developments. It is not the purpose of this paper to question or evaluate the philosophy and practice of the religion of Scientology. Instead, by treating Scientology as a "black box" and focusing on the outcomes of the application of Scientology practice, unnecessary controversy can be avoided. The analysis of outcomes would

---

seem to be a relatively straightforward scientific endeavor (although non-trivial), comparable to the studies of psychotherapeutic outcomes, or educational evaluation. This paper will show that previous studies of scientology outcomes are inconclusive, and will describe a possible approach to new research.

In 1950, responding to popular demand, Hubbard initiated an attempt to validate the outcomes of dianetic therapy. Hubbard realized that case histories and testimonials were subjective and susceptible to bias, and instead required that cases be evaluated with objective tests of intelligence and personality, and in some cases, medical examinations, before and after dianetic therapy. The preliminary scientific validation of dianetic processing conducted in 1950 by the Hubbard Dianetic Research Foundation (Ibanez, Southon, Southon, & Benton, 1951) used a variety of intelligence and personality tests given before and after processing. A few individual case histories included Minnesota Multiphasic Personality Inventory (MMPI) profiles, Wechsler intelligence tests, and medical examinations, but the largest sample consisted of eighty-eight students on the professional dianetic practitioner's course, who were administered a battery of tests at the beginning of their training in August 1950, and again in mid-September. The battery consisted of the California Test of Mental Maturity Short Form (CTMM), the Johnson Temperament Analysis, the California Test of Personality, and the Mental Health Analysis. During their training, they both received and practiced dianetic auditing on other students. The CTMM showed a mean gain of 7.9 IQ points. According to Jensen (1980, p. 590) "Retesting of naïve subjects on the identical test, after a short interval, shows gains of about 2 to 8 IQ points for various tests, averaging about 5 IQ points." Since no alternate forms of the CTMM were available, and many of the middle-aged examinees had probably never taken tests with machine-scorable answer sheets, it is possible that the Dianetic Course IQ gains may be entirely due to practice effects. All three personality tests also showed improvements. There is no way to know from the report whether these personality test changes reflect temporary mood swings or stable personality changes, because there were no follow-up studies. The study's authors considered that their work was only preliminary, partly because it did not include a control group. Their plans for a more ambitious study were interrupted by the bankruptcy of the Hubbard Dianetic Research Foundation in early 1951, and no comparable effort was ever mounted again.

In his doctoral dissertation, Fischer (1953) failed to find significant changes in intelligence or personality from "dianetic auditing" in a controlled experiment with three groups of 12 subjects who received zero, 1-15, and 16-36 hours of treatment, respectively. Unfortunately, Fischer conducted his research during a time when the bankrupt Hubbard Dianetic Foundation was being operated in receivership, after Hubbard had resigned and the courses were being taught by psychologists and a medical doctor who were more interested in promoting their own techniques than anything resembling dianetic auditing. The subjects in Fischer's experiment received treatment twice a week in one or two hour sessions, a schedule that is not unusual in psychological practice, but which would make it impossible to use dianetics to treat major traumatic incidents, which can require five or more consecutive hours for complete desensitization. Thus, it is almost certain that Fischer's subjects did not receive dianetic auditing, as he believed.

Another problem with Fischer's experiment is that the small sample sizes were not likely to produce significant results. In order to have an 80% chance of obtaining a significant result for a medium effect size, 52 subjects in each group would have been required (Cohen, 1988).

With the cooperation of the Church of Scientology, Ross (1988) mailed attitude and demographic questionnaires and personality tests to 100 long term members of Scientology who ranged from 2.5 years to 30 years in the movement, receiving 48 answers back. He correlated a number of variables with length of time in Scientology, partialled for age, finding several significant correlations with favorable traits, and none with unfavorable traits. Since correlation does not imply causation, it was not clear whether Scientology membership improves personality or whether initially strong characters remain longest. However, the findings were not consistent with anti-cultists' claims that members were unhappy or unstable prior to joining, or that they were damaged by their Scientology experience.

Although formal efforts to validate dianetic or Scientology outcomes were abandoned, before-and-after testing continued, with evidence accumulating on tens of thousands of cases. Because of its large sample size, this data base is a priceless resource for studies of sub-populations of special cases receiving different types of treatment. It might be possible to "mine" this data base to arrive at scientific conclusions by comparing cases during the times they received auditing only, training only, or no services during an absence. Interesting questions might be answered concerning what types of people benefit from auditing and from which of a variety of techniques. The success of such an endeavor depends critically on the quality of the tests administered to the Scientology clients. The remainder of this paper is designed to examine this issue.

Among the tests used in the Ibanez et al. (1951) study was the *Johnson Temperament Analysis* (JTA) (Johnson, 1944b). Meltzer (1949) and Ellis (1953) unfavorably reviewed the JTA for its inadequate information on validity, reliability and usefulness. The JTA was extensively revised in 1963 and renamed the *Taylor-Johnson Temperament Analysis* (T-JTA). Recent reviews are generally favorable (Axford, 2005; Boyle, 2005), although critical of the fact that test validity information is based entirely on correlations with other tests.

The JTA was last used by scientologists in a study by Julia Lewis of Hubbard Guidance Center auditing in Phoenix (Lewis, 1955a). At the same time, Lewis announced a program to develop "scientometric" tests to replace the psychometric ones previously used. Later, Lewis (personal communication, 1964) said that Hubbard had come to her saying that he had been cut off from access to psychological tests[1]. At his request, Lewis (1995b) developed the *American Personality Analysis* (APA) to replace the Johnson Temperament Analysis. Lewis said that the APA was "based on" the Johnson, but otherwise did not clarify the relationship of the APA to the Johnson. The APA was published in 1955 and copyrighted by Julia M. Lewis, BA, BSCN, DSCN, Director of the Hubbard Guidance Center, Washington, D. C.

According to Raymond Kemp (1995, p. 21), Hubbard brought up the matter of the APA to him, stating that "the test results did not seem to fit the observed results on students and pcs in London." Hubbard "asked me whether it would be possible to write a test that was more general in nature, and would enable him to see in the test what he was looking for. He also wanted it to be in the same general format as the APA and if possible to have both tests interchangeable in the matter of what he wanted to see as information."

---

[1] About this time, test publishers began to require that their customers be members of the American Psychological Association.

In 1959, the Oxford Capacity Analysis was published (Kemp, 1959) The first edition was copyrighted by Kemp, but later editions are copyrighted in L. Ron Hubbard's name.[2] About 75% of the OCA manual is paraphrased or copied verbatim from the APA manual (Lewis, 1955b); the rest contains additional material. There is no mention of Raymond Kemp or Julia Lewis in the manual, although the OCA has many similarities to the earlier tests.

Probably because the copyrights were transferred to L. Ron Hubbard, there is a widespread misconception that Hubbard was the author of the APA/OCA. Nothing could be further from the truth. Only a handful of Hubbard's writings or lectures even mention the OCA, and all of these have to do with the interpretation and use of the OCA. Hubbard himself explicitly disavowed any expertise in psychometrics and relied on the advice of others (Hubbard, September 23, 1950/1994, p. 643).

Personality testing has been used for three purposes by the Church of Scientology. The earliest use was to demonstrate the effects of Scientology auditing by comparing scores before and after auditing. The second use was to serve as an aid to case supervision. The third use was to serve as a tool for interesting newly contacted public in aspects of their personality that could be improved if they enrolled in Scientology services.

Sir John Foster (1971) conducted an official British government inquiry into Scientology. As part of that study, three eminent psychologists appointed by the British Psychological Society investigated the OCA, and came up with extremely negative conclusions. Unaware of the OCA's close relationship to the JTA, they concluded that "the Oxford Capacity Analysis is not a genuine personality test; certainly the results as presented bear no relation to any known methods of assessing personality" (Foster, 1971, p. 76). They experimented with entering random responses on the answer sheet and found that several of the resulting OCA scores were very low. They asserted that a "random method of response … would be expected to produce scores close to the mean of scores obtained during the standardizing of the test" (Foster, 1971, p. 76). It is a mystery how anyone could believe this, when any schoolchild soon learns that randomly responding to the questions on an achievement examination will result in failing the exam, and not a score near the mean. Nevertheless, they claimed that random responses to all known personality inventories would result in flat profiles near the mean, and therefore concluded that the OCA must be unreliable. However, one of the most thoroughly validated tests in existence, the MMPI, has extremely abnormal profiles when the answer sheets are marked randomly, all True, or all False (Duckworth & Anderson, 1995, pp. 44-63), so the Foster report is surely wrong on this point. Apparently unaware of the first two uses of testing in Scientology, the experts concentrated their criticism exclusively on the use of the OCA for proselytizing. They justifiably faulted the test administrators for ignorance of psychometrics and OCA test construction, thus risking giving incorrect interpretations to the examinees.

---

[2] "Copyright London by Raymond Kemp, Ph.D. and the Hubbard Guidance Centre." A new edition, called *The Standard Oxford Capacity Analysis Question Booklet* (New Era Publications, 1987) has a copyright 1978 by L. Ron Hubbard and 1987 by L. Ron Hubbard Library. The test manual is copyrighted 1955, 1968, 1979, 1980, 1986 by L. Ron Hubbard, but the author is unspecified. (The 1955 copyright date is anomalous, because the test itself was not published until 1959.)

PURPOSE

The paper is devoted to the question of whether the existing data base of several thousand case histories containing of OCA test scores is of sufficient quality that it could be used for scientific outcome studies. This study was intended to facilitate scientific research on Scientology outcomes by determining the characteristics of Scientology personality tests, their defects, and possible remedies. While it has been generally known that the APA and OCA were "based on" the Johnson Temperament Analysis, the exact relationship has been something of a mystery. The present study was designed to describe the relationship in some detail in order to discover possible deficiencies and avenues of improvement of the tests. Although this critique of the Scientology tests will be severe, it is intended to be helpful and constructive, rather than hostile. This paper is written for religion scholars, sociologists, psychologists, and at least a few scientologists interested in outcome research who will have to work together across disciplines if any of the suggestions in this paper are to be implemented.

METHOD

Copies were obtained and compared of the test booklets, test scoring keys, and test manual for the Johnson Temperament Analysis (Johnson, 1944a, 1944b, 1944c), American Personality Analysis (Lewis, 1955b, 1955c, 1955d), and Oxford Capacity Analysis (Kemp, 1959); *The Manual of the Oxford Capacity Analysis* (New Era Publications, 1986); *The Standard Oxford Capacity Analysis Question Booklet* (New Era Publications, 1987); *The Standard Oxford Capacity Analysis Marking Sheet* (New Era Publications, 1981); *Standard OCA Percentile Charts* (New Era Publications, 1978). First, the names of the scales in the JTA were compared with those of the other tests. Then, the items within each scale of a test were matched for similar meanings in the corresponding scales of the other tests. For items that matched, the scoring weights were compared. Finally, several mathematical transformations were made on the test norms for the JTA in an attempt to predict the norms for the APA/OCA. Several hypotheses were examined:

Hypothesis 1. The items in the APA/OCA are paraphrased from the JTA.

Hypothesis 2. The scoring weights of the APA/OCA are identical to, or simple transformations of the weights of the JTA.

Hypothesis 3. The test norms of the APA/OCA are linear transformations of those of the JTA.

RESULTS

The Scales

The Johnson Temperament Analysis measured traits in terms of opposite tendencies, e.g. *Active* vs. *Quiet*. The APA and OCA carried on this dichotomous description of personality traits. They measured the same traits as the JTA, but each named them differently. The APA added a new scale with new items, called *Serene* vs. *Manic (cycloid).* Four of the nine traits measured by the JTA are negative, i.e., high scores are undesirable. These are reversed in the APA and OCA, so that high scores are always desirable.

Table 1 compares the scale names for the three tests. No one really knows how accurately the names describe what the tests actually measure. Notice that in the JTA, *Appreciative* is the opposite of *Critical*, whereas in the OCA, it is opposite *Lack of Accord* . The JTA *Appreciative* has been renamed *Correct Estimation* in the OCA, and the JTA *Sympathetic* has been renamed *Appreciative* in the OCA.

Table 1

Trait Scale Names for the Johnson Temperament Analysis, American Personality Analysis, and the Oxford Capacity Analysis

JTA		APA		OCA	
High	Low	High	Low	High	Low
Self Mastery	Impulsive	Oriented & Self Controlled	Impulsive & Dispersed	Stable	Unstable (Dispersed))
Depressive	Gay-hearted	Happy	Depressed & Unhappy	Happy	Depressed
Nervous	Composed	Composed	Scattered & Nervous	Composed	Nervous
-------------	-------------	Serene	Manic (Cycloid)	Certainty	Uncertainty
Active	Quiet	In Motion	Stopped	Active	Inactive
Aggressive	Submissive	Overt & Aggressive	Inhibited & Submissive	Aggressive	Inhibited
Subjective	Objective	Being Cause & Objective	Being Effect & Subjective	Responsible (Causative)	Irresponsible
Critical	Appreciative	Correct Estimation & Appreciative	Critical	Correct Estimation	Critical
Sympathetic	"Hard-boiled"	Empathy & Sympathetic	Lack of Accord	Appreciative	Lack of Accord
Cordial	Cold	Expressive of Affinity	Withdrawn & Cold	Comm. Level	Withdrawn

The Items (Questions)

All of the items of the JTA, 1959 OCA, and 1987 OCA were aligned in a spreadsheet, along with the weights given for the item responses. Unfortunately, this spreadsheet cannot be published without violating copyrights for all of these tests. In what follows, all item wordings are artificially constructed examples designed to illustrate a point. None of them are actual item wordings.

Almost all of the items in the APA and OCA are paraphrases of the original JTA test. I will distinguish and illustrate different levels of paraphrase in Table 2.

Table 2
Levels of Item Paraphrase

Paraphrase Level	Example
JTA Item	Does S have a fear of black cats?
Level 0 paraphrase	Do you have a fear of black cats?
Level 1 paraphrase	Are you afraid of black cats?
Level 2 paraphrase	Do black cats frighten you?
Level 3 paraphrase	When you see a black cat in your area, do you get nervous?
Level 4 paraphrase	Do you feel a superstitious dread when a black cat crosses your path, even though you know it is silly?

The JTA was originally designed as a questionnaire that a friend or relative could fill out about the person; hence all JTA items start out with "Does S …" When the JTA was filled out by the examinee, s/he was expected to translate this into "Do you …" Thus, the Level 0 paraphrase is a minimal change.

The level 1 paraphrase uses a slight change of wording but preserves the grammatical structure with the same subject and prepositional phrase "of black cats."

The level 2 paraphrase makes "black cats" the subject and "you" the object of a new transitive verb, "frighten." Thus the grammatical structure of the sentence, as well as the wording, has changed, while preserving the meaning.

The level 3 paraphrase introduces more extensive grammatical changes and a slight difference in meaning.

The level 4 paraphrase substantially alters the meaning as well as the grammatical structure, but is addressed to the same topic of black cats and fears.

A level 5 paraphrase would be an almost completely different item.

Table 3 shows the number of JTA items that were paraphrased at different levels in the APA and OCA, based on counts I made using my own classification system. The table confirms Hypothesis 1.

In addition, the APA and OCA contain a tenth scale D, called *Serene* on the APA and *Certainty* on the OCA. Two of the items (included in Table 3) are level 0 to 2 paraphrases of items that appear on the JTA questionnaire but were not scored on any JTA scale. The other 18 items appear to be original.

Table 3
Number of JTA Items Paraphrased in the APA and OCA at Various Levels

Paraphrase Level	1955 APA	1987 OCA
0	32	13
1	90	55
2	48	67
3	7	30
4		10
5		2
Total	177	177

The OCA was clearly based on the APA, and contained a number of identical items. Some of the terms are peculiarly British words or idioms that would be hard for Americans to understand. Other questions appear to be improvements on the APA[3].

How the Tests Are Scored

Each item has three *response weights* associated with it. Each response weight is the number of points the examinee receives by making that response. *Notation*: In this paper, the three response weights will be separated by commas. For example, an item with the weights (7,5,1) will give the examinee 7 points if s/he answers YES (+), 5 points if s/he answers MAYBE, uncertain or sometimes (m), and 1 point if s/he answers NO (-). An examinee's *raw score* is defined as the sum of the weights for the item responses s/he made.

The OCA (The Standard Oxford Capacity Analysis marking sheet, 1981) has exactly the same item response weights as the APA, with the exception of one item that should have been reversed on the APA. The APA's weights (Lewis, 1955d) are closely related to the JTA's (Johnson, 1944c), and in most cases are identical. Table 4 shows the relationship between the APA and the JTA scoring.

---

[3] This copy of the 1959 OCA was evidently a rough draft, with 13 misspelled words, not including 6 English words with U.K. spellings. None of the questions end with a question mark. The errors were corrected in the 1987 Standard Oxford Capacity Analysis, which is closer to the original APA than is the 1959 OCA.

All of the scales on the APA have 20 items in them. The JTA has 20 items in each of six scales, 19 in one scale, and 18 in two other scales. Thus the APA added 5 new items, each with their own response weights, to the JTA scales. Also, one item on the *Correct Estimation (Appreciative)* scale was completely different. In addition, 5 items had slightly different weights, possibly due to clerical errors. These changes are shown in the fourth column of Table 4.

Table 4
Scoring Weight Changes from the Johnson Temperament Analysis to the American Personality Analysis

APA Scale Name	APA Scale	JTA Scale	Weight Changes	Added Items
Oriented	A	I	1 change: 6,3,2 →6,3,3	
Happy	B	-B	No change	
Composed	C	-A	1 change: 6,2,2 →2,6,6	1 added: 6,4,4
Serene	D		(New APA Scale)	
In Motion	E	C	2 changes: 5,4,3 →6,4,3 2,4,6 →3,4,5	2 added: 0,1,5 0,1,5
Aggressive	F	G	1 change: 6,3,3 →6,6,3	2 added: 6,4,2 0,3,5
Objective	G	-F	No change	
Correct Estimation	H	-H	1 different question: 6,4,2 →2,5,6	
Empathy	I	E	No change	
Affinity	J	D	No change	

The APA scales were arranged in a different order and had different letters assigned to them than the JTA. The middle two columns of Table 4 show the relationship. As previously mentioned, four of the JTA scales measured negative traits. Where these scales were reversed in the APA, the JTA scale letters are preceded with a minus sign. The main point of Table 4 is not the scoring weight changes, but the fact that there are so few of them. Almost all of the weights are the same (or reversals), confirming Hypothesis 2.

Reversed scale scoring.

In order to reverse the scoring of the JTA items for the reversed scales, Lewis interchanged the first and third weights while keeping the middle weight unchanged. I will refer to this as the *interchange reversal* method. For example, if the item "Are you afraid of black cats?" was scored (7,5,1) on the JTA *Nervous* scale, the same item would be scored (1,5,7) on the APA *Composed* scale.

Here Lewis made a major error. In the JTA, a middle response is 2 points away from a full *Nervous* response and 4 points away from a *Composed* response. In the APA reversal, the middle response is 4 points away from a full *Nervous* response and 2 points away from the *Composed* response.

Occasionally, instead of reversing the weights, Lewis reversed the meaning of the item, e.g., "Are you unafraid of cats?" and kept the weights at (7,5,1). This runs into the same problem as before, namely the middle response is 2 points away from the *Composed* response, not 4 points as in the original JTA version.

Interchange reversal suffers from three undesirable properties: (1) Two examinees with the same original scores can have different reversed scores. (2) Two examinees with different scores can have the same reversed score. (3) The rank order of examinees is not necessarily preserved. For example, one examinee might be more depressed than another on the JTA scale but happier than the other on the APA/OCA scale. The proofs are easy[4].

The correct reversal method, which will be called *subtractive reversal*, would be to subtract each weight from the sum of the high and low weights. The reversal of (1,5,7) would be ((8–7), (8–5), (8–1)) = (1,3,7) for the APA item response weights. It can be easily shown that there is a simple one-to-one linear functional relationship between the original scores, S, and the subtractive reversed scores, S', namely, $S' = K - S$, where K is a constant for this scale equal to the sum of the YES scoring weights and the NO scoring weights over all of the scale items[5].

Test Scale Norms

When a person answers the questions on an aptitude test or personality questionnaire, his/her responses to a certain set of items are weighted appropriately and added up into a raw score. To be interpreted, the raw score must be compared with the scores of other people in

---

[4] Let two examinees, A, and B, have identical responses except on two items of a scale, where A answers YES to the first question and MAYBE to the second question while B responds MAYBE to the first question and YES to the second question. For assertion (1) let the weights be (1,2,3) and (1,2,4). Then A and B both have scores of 3, but their reversed scores are 5 and 6 respectively. For assertion (2), let the weights be (3,2,1) and (4,2,1). A and B have scores of 5 and 6 but identical reversed scores. For assertion (3), let the weights be (1,2,5) and (1,3,4). A and B score 4 and 3 but 8 and 6 on the reversed scale.

[5] Proof: Let W_{ij} be the weight of response j to item i. Let $X_{ij} = 1$ if the examinee answers with response j to item i and 0 otherwise. Then the scale score, $S = \sum_i \sum_{j=1}^{3} W_{ij} X_{ij}$. The subtractive reversed score,

$$S' = \sum_i \sum_{j=1}^{3} (W_{i1} + W_{i3} - W_{ij}) X_{ij} = \sum_i W_{i1} + \sum_i W_{i3} - S .$$

his/her reference group. For example, if he is a college senior, his scores are best compared with those of other male college seniors, and not with high school girls, army recruits, or inmates of mental institutions. The process of tabulating the raw score statistics for appropriate reference groups is called *norming*.

The test manual for the Johnson Temperament Analysis (Johnson, 1944a) gives norms for four different reference groups: Men, Women, Boys, and Girls. The sample sizes were 200 in each group, and each group was described in some detail. The APA and OCA give norms for the same four reference groups, without describing the groups or specifying their sample sizes.

Once norming has been accomplished, it is usual to convert raw scores into some kind of norm-referenced score, such as a standard score or a percentile score.

Standard scores have a specified mean and standard deviation in the reference group. For example, if they had a mean of zero and a standard deviation of 40, they would vary from about -100 to +100. They have a simple, straight-line relationship to the raw scores.

A *percentile* is defined as a value on a scale of one hundred that indicates the percent of a distribution that is equal to or below it. For example, a percentile score of 95 is a score equal to or better than 95 percent of the scores in a reference population. Obviously, the lowest possible percentile score is zero, and the highest possible score is one hundred. When plotted on a graph against raw scores, the percentiles follow a smoothly rising curve that is nearly horizontal at the beginning and at the end and steepest in the middle.

For example, the manual for the JTA shows the percentiles corresponding to raw scores of Trait B *Depressed*, for a group of 200 women tested before 1944, possibly before 1941. The sample was made up of women over 18 years old in business college, night courses of high school level, women clients at the American Institute of Family Relations in Los Angeles who had come in for pre-marital instruction only, a few university students and others from women's clubs.

The manual for the American Personality Analysis (Lewis, 1955b) shows the corresponding norms of women for Trait B (*Happy – Depressed*) of the APA. The most striking feature of the table is the presence of negative percentiles! There is no hint in the test manual of what such a concept might mean. One is forced to conclude that the author of the test (Julia Lewis) had a misunderstanding of the term "percentile," and that everyone who has used the test has had a similar misunderstanding.

To avoid further confusion, we shall use the term *bicentile* to describe the scale from -100 to +100 in the APA. The bicentile score, *B*, is defined as
$B = 2*(P – 50)$, or -2*(P – 50) in a reversed scale, where *P* is the JTA percentile. Thus, a zero bicentile score represents the median score of the population.

The APA test manual presents no information on the characteristics of the reference group that was used to norm the test. The sample size, ages of examinees, their education level, location, and other relevant information are all omitted.

There is good reason to believe that no data were collected to construct the norms. Our Hypothesis 3 is that the norms of the APA were derived from those of the JTA by mathematical formula. If the scales are not reversed in the APA from those in the JTA, then the raw scores in

the tables are the same in both tests. If the scales are reversed by interchange reversal, the lack of functional relationship between the original scores and reversed scores makes it impossible to correctly translate the JTA norms into norms for the reversed scores. Lewis must have improvised some ad hoc (but incorrect) method, which we can only guess to be

$S'_{APA} = K - S_{JTA}$, where K is an empirically derived constant for this particular scale and reference group. Possibly, Lewis could have established a value of K by scoring one examinee both with JTA weights and with APA weights and adding the two scores, but the result would have been different for a different examinee. In this paper, each scale K that Lewis used was estimated by matching the JTA and APA at a single point near the median of that scale.

Figure 1 compares the transformed JTA percentiles and raw scores with the APA bicentiles for women on the reversed scale *Happy*.
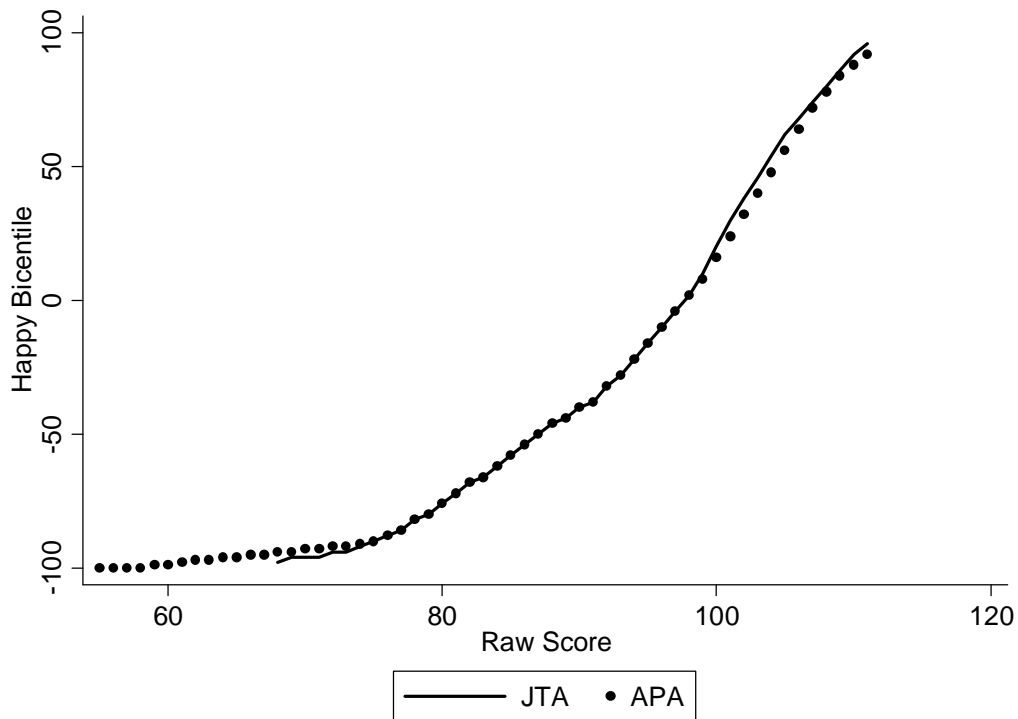

Figure 1. JTA and APA Bicentiles for the *Happy* Trait for 200 Women.

Hypothesis 3 is only partially confirmed. In the region from -90 bicentile to +2 bicentile, the formula predicting the APA from the JTA works perfectly. Above and below those levels, the absolute values of the APA bicentiles are slightly less than what the formula predicts. Analysis of the other scales produces similar results: there is a region where a formula works perfectly, above or below which the APA bicentiles have smaller absolute values than would be expected.

It can be conjectured that Lewis was dissatisfied with a basic property of percentile transformations: that at low and high scores, large changes in the raw scores produce only small changes in the percentile scores, since the latter are near their limits already. Her solution for avoiding this was to make the percentile scores reach their limits more slowly by replacing the curve in some places where it is steep by a straighter line with a shallower slope. A better solution would have been to abandon percentile scores and use standard scores, which are linear functions of the raw scores.

There are three scales on the APA that have the same number of items as on the JTA and which were not reversed: A. *Oriented*, I. *Empathy*, and J. *Expressive of Affinity*. The bicentiles computed from the JTA exactly match those of the APA throughout most of the range, except at the upper and lower extremes. There are a very few exceptions in the middle ranges. For example, on the *Expressive of Affinity* scale for High School Girls, the APA bicentile at raw score 78 is lower than at 77. That is, the bicentile score decreases when the raw score increases at that point. This is clearly a typographical error.

Two other unintentional errors in the APA norms interfered with perfect prediction from the JTA norms: (a) On the *Oriented* scale for Men, the APA bicentile for a raw score of 90 is repeated for 91, and for 91 on up, the correspondence with the JTA is off by one. (b) On the *Happy* scale for High School Boys, a raw score of 89 is omitted from the APA/OCA table, and at that point the correspondence shifts.

Most of the other scales of the APA can be fitted to a transformed JTA by a similar procedure. However, there are two scales that had 18 items in the JTA and 20 in the APA; these are E. *In Motion*, and F. *Aggressive.* They fit well for the negative bicentiles, but depart radically for the positive ones. Figure 2 illustrates the relationship for the *In Motion* scales. The graphs for the *Aggressive* scales are similar.
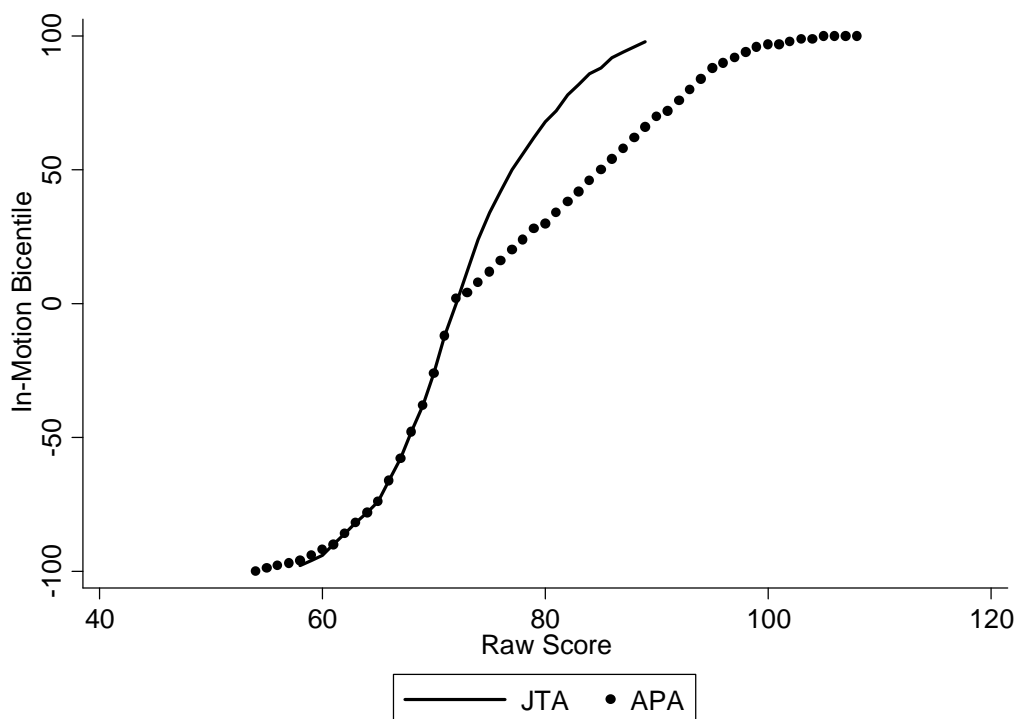
Figure 2. JTA and APA Bicentiles for the APA *In-Motion* Trait for 200 Women.

In all cases, the positive APA bicentile is well below the computed JTA bicentile. These could not possibly describe any real statistical observations. Adding two items would change not only the maximum bicentiles, but would affect the entire range. In particular, the zero bicentile would be shifted to a larger raw score. But unfortunately, there is no correct procedure for computing the bicentiles for a different scale with two more items in it, other than to collect data by administering the expanded tests to reference groups.

Possibly Lewis, faced with an intractable problem, simply decided to invent numbers to put down in the tables for these two traits.

Finally, the APA bicentile curve for the D. *Serene (Certainty)* scale that was added to the APA has a curve unlike any of the others. It is an irregular line with little curvature. It could not have been based on any tabulated observations; probably it was merely invented.
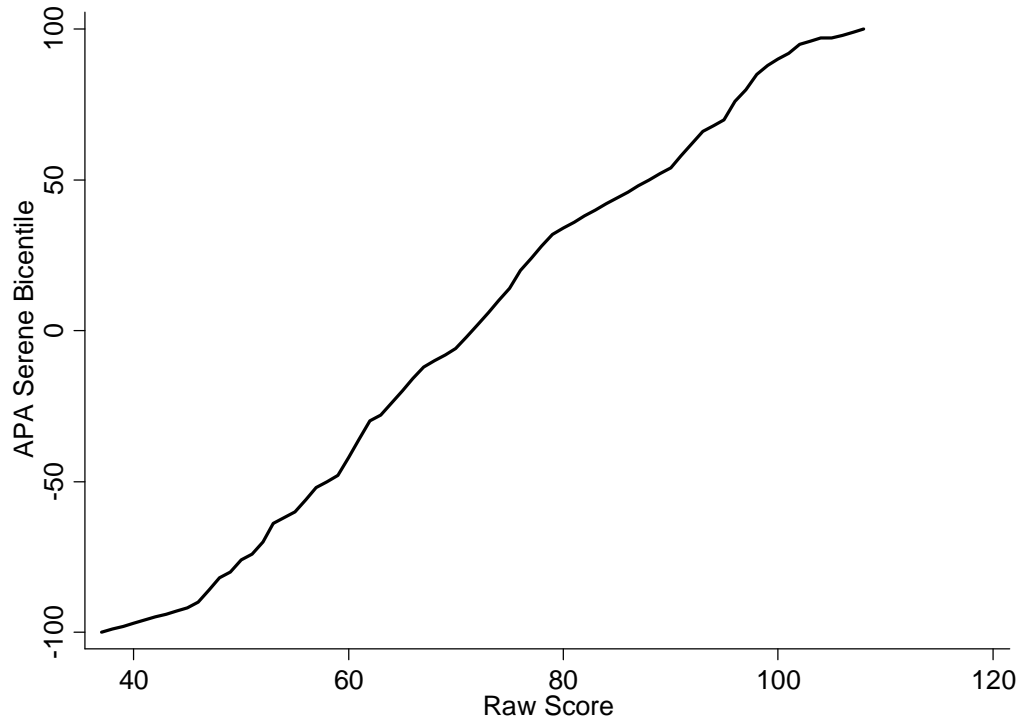
Figure 3. APA Bicentiles for the *Serene* Trait for Women.

DISCUSSION

The OCA was derived from the Johnson Temperament Analysis (JTA) without accurately duplicating any of its psychometric characteristics of item weights, test norms, reliability, and validity. The paraphrased questions in the OCA are unlikely to have the same response frequencies[6] or correlations with their scale scores as their JTA counterparts. Also, the OCA and the JTA have obsolete test norms and questions. These facts pose a dilemma for the Church of Scientology, because changing the OCA risks damaging the effectiveness of case supervision methods, which rely on the OCA (Hubbard, 1991).

Despite the lack of a mathematical one-to-one relationship between the OCA and the JTA, it seems likely that there is a large statistical correlation between the OCA and the JTA, as there is between the JTA and its revision, the Taylor-Johnson Temperament Analysis (T-JTA). Verification of this conjecture may offer a solution to the problem posed above. Verifying the correlations between the OCA, JTA, and T-JTA on a sample of a few hundred cases would

---

[6] Schuman and Presser (1996) reported data on two versions of the same question in an opinion survey. In one survey, 44% of the respondents said "No" when asked, "Do you think the United States should allow speeches against democracy?" But when asked, "Do you think the United States should forbid speeches against democracy?" only 28% said they should be forbidden.

support the reliability and validity of the OCA, because the T-JTA itself has established validity and reliability.

Additionally, the OCA should be validated against empirical criteria, such as ratings or adjective check lists filled out by individuals who know the examinee well, such as roommates, co-workers, or family members.

Existing examinee answer sheets could easily be used to determine OCA item statistics, internal consistency reliability , and scale intercorrelations.

It would be valuable to measure the degree to which changes in a person's test scores reflect changes in his/her real-world behavior. The validities of changes in scale scores could be determined by a study in which individuals finishing Scientology services rated themselves on how much change they experienced in several areas of their lives, and these ratings were correlated with the OCA changes.

Establishing some degree of validity for the OCA would make the large data base of OCA test results in the Church of Scientology case history archives relevant and useable for outcome research. This is the main point of this paper.

For future use, the defects of the OCA could be corrected, and its validity improved by careful construction of a wholly new test designed to measure the same traits as the OCA, but with completely original questions, simplified scoring, norms based on actual data, and standard scores instead of bicentiles. The new test should be equated to the old test so that scores are, on the average, approximately equal for the same individuals. An initially large item pool could be refined so as to improve validity, and the final version field tested and validated against peer ratings, not just other test scores. Even if the new test seems technically perfect, extreme caution should be used to ensure that the new test does not reduce the quality of case supervision.

REFERENCES

Axford, S. N. (2005). [Review of the test Taylor-Johnson Temperament Analysis]. In R. A. Spies & B. S. Plake (Eds.), *The Sixteenth Mental Measurements Yearbook* [Electronic version]. Available from *http://www.unl.edu/buros/*

Boyle, G. J. (2005). [Review of the test Taylor-Johnson Temperament Analysis]. In R. A. Spies & B. S. Plake (Eds.), *The Sixteenth Mental Measurements Yearbook* [Electronic version]. Available from *http://www.unl.edu/buros/*

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences,* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Duckworth, J. C., & Anderson, W. P. (1995). *MMPI & MMPI-2: Interpretation manual for counselors and clinicians,* (4th ed). Levittown, PA: Taylor & Francis.

Ellis, A. (1953). [Review of the test Johnson Temperament Analysis]. In O. K. Buros (Ed.), *The Fourth Mental Measurements Yearbook* (pp. 62-63). Highland Park, NJ: Gryphon Press.

Fischer, H. J. (1953, September). *Dianetic Therapy: An experimental evaluation. A statistical analysis of the effect of dianetic therapy as measured by group tests of intelligence, mathematics and personality*. Unpublished doctoral dissertation, School of Education of New York University.

Foster, J. (1971). *Enquiry into the practice and effects of Scientology*. London: Her Majesty's Stationery Office. Retrieved from http://www.cs.cmu.edu/~dst/Cowen/audit/fosthome.html

Hubbard, L. R. (1950). *Dianetics: The modern science of mental health*. New York: Hermitage House.

Hubbard, L. R. (1975). *Dianetics and Scientology technical dictionary*. Los Angeles: Publications Organization.

Hubbard, L. R. (1991). The Expanded Dianetics Series. In *The Technical Bulletins of Dianetics and Scientology. Subject Volume 3*: (pp. 129-459). Los Angeles: Bridge Publications.

Hubbard, L. R. (1994). What dianetics can do. In *Research and discovery series: Vol. 4:* (pp. 635-653). Los Angeles: Bridge Publications. (Original lecture September 23, 1950).

Ibanez, D., Southon, G., Southon, P., & Benton, P. (1951). Dianetic processing: A brief survey of research projects and preliminary results. Elizabeth, NJ: Hubbard Dianetic Research Foundation.

Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.

Johnson, R. H. (1944a). *Manual of the Johnson Temperament Analysis*. Los Angeles: California Test Bureau. Retrieved from Test Collection, Carl Campbell Brigham Library, Educational Testing Service, Princeton, NJ.

Johnson, R. H. (1944b). *Johnson Temperament Analysis, Form A*. Los Angeles: California Test Bureau. Retrieved from Test Collection, Carl Campbell Brigham Library, Educational Testing Service, Princeton, NJ.

Johnson, R. H. (1944c). *IBM Form 805 weighted scoring keys for traits A-J, Johnson Temperament Analysis, Form A*. Los Angeles: California Test Bureau. Retrieved from Test Collection, Carl Campbell Brigham Library, Educational Testing Service, Princeton, NJ.

Kemp, R. (1959). *Oxford Capacity Analysis*. London: Hubbard Guidance Center.

Kemp, R. (1995). Kemp's column. Putting it to the test. *International Viewpoints*, *22*, 20-26. ISSN 0905-9725.

Lewis, J. M. (1955a). Analysis of processing results. *Ability, 2* (Major), 12-14.

Lewis, J. M. (1955b). *Manual of the American Personality Analysis*. Washington, D.C.: Distribution Center.

Lewis, J. M. (1955c). *American Personality Analysis test booklet*. Washington, D.C.: Distribution Center.

Lewis, J. M. (1955d). *Scoring keys for the American Personality Analysis.* Washington, D.C.: Distribution Center.

Meltzer, H. (1949). [Review of the test Johnson Temperament Analysis]. In O. K. Buros (Ed.), *The Third Mental Measurements Yearbook* (p. 99). Piscataway, NJ: Rutgers University Press.

New Era Publications. (1978). *Standard OCA percentile charts*. Copenhagen: New Era Publications.

New Era Publications. *(1981). The standard Oxford Capacity Analysis marking sheet.* Copenhagen: New Era Publications.

New Era Publications. (1986). *The manual of the Oxford Capacity Analysis* . Copenhagen: New Era Publications.

New Era Publications. (1987). *The standard Oxford Capacity Analysis question booklet.* Copenhagen: New Era Publications.

Ross, M. W. (1988). Personality changes in Scientologists: Effects of membership. *Journal for the Scientific Study of Religion, 27*: 630-636.

Schuman, Howard and Stanley Presser (1996). *Questions and Answers in Attitude Surveys.* Thousand Oaks, CA: Sage Publications.