CareerCruising

MatchMaker Reliability and Validity Analysis

Career Cruising MatchMaker Analysis
Reliability and Validity Analysis

CECS contracted with Career Cruising to conduct psychometric analyses of the MatchMaker instrument used in the CareerCruising online career guidance system. Results of factor analyses of the core 39 items and the expanded item set were previously reported as were results describing the internal consistency reliability of the resulting factors from both series of analyses.

In this report we describe results from a series of analyses designed to describe (1) the test-retest reliability of the MatchMaker items and factors (including analyses from the core 39 item set and extended item set) and (2) the concurrent construct validity of scales derived in previous analysis.

Test-retest Reliability

Test-retest reliability describes the stability of examinees' responses to items or combinations of items (scales) over a fixed period of time. It is appropriate to describe the test-retest reliability of item or scale scores for instruments that measure underlying constructs that are thought to be stable over time. It is generally not recommended to use this type of reliability analysis to describe underlying constructs that are transient, state-dependent, or have strong proximal developmental influences.

There is ample empirical evidence describing the stability of career interests. Recent meta-analyses of interest measures suggest that career interests begin to stabilize in early adolescence and become quite stable in young adulthood. Thus, an analysis of MatchMaker interest stability is not only appropriate but will assist CareerCrusing to establish the psychometric foundations of its flagship inventory.

Construct Validity

Construct validity is a complex concept. In its broadest sense, construct validity answers the question, "are we measuring what we think we are measuring?" When applied to an interest measure such as MatchMaker, the question becomes "how confident are we that we are measuring the full universe of career-related interests?" One of the most common ways of establishing the construct validity of an instrument is to compare scores on that instrument to scores on another instrument of that same construct in a single sample. This procedure was used to explore the construct validity of the MatchMaker core and extended items set. Strong relationships between scores on the MatchMaker and like-scores on another interest inventory would be evidence in support of the construct validity of the MatchMaker item sets and further support the use of this instrument with adolescents and adults in prescribing possible educational and career paths.

Methodology
Test-Retest Reliability

CareerCruising provided CECS with 4 separate files that included samples of students who had completed MatchMaker items at two points in time. In two samples, examinees completed the MatchMaker items sets between 7 and 21 days apart and in two additional samples, examinees

completed the MatchMaker item set between 90 and 120 days apart. The table below describes the range and average test-retest latencies in the four samples.

	MinTime Days	MaxTime Days	Average Time	N
HS 7-21	7	21	12.28	36820
HS 90-120	90	120	105.57	12150
College 7-21	7	21	12.08	1047
College 90-120	90	120	104.53	97

It should be noted that the ability to generalize test-retest reliability coefficients is partly dependent on sample size. Thus considerably more confidence can be placed on results from the high school sample compared to the adult sample. Results from the 90-120 day test-retest analysis in college students should be considered preliminary.

Data were cleaned to remove students who had homogenous response patterns (e.g., all 1s or all 5s) and pairwise deletion methods were used in all analyses. The standard convention for describing test-retest reliability is the pearson product-moment correlation coefficient. Correlations were conducted separately at the item level (e.g., correlation between item 1 at time 1 and item 1 at time 2) and at the factor level. Factors (scales) used in this study were derived from previous analysis describing the Holland-like factor structure inherent in the MatchMaker item sets. Further, two factor (scale) solutions were used in this study. The first solution included the six factors extracted using only the 39 items in the core MatchMaker item pool whereas the second solution used the six factors extracted in previous analysis using the best 7 performing items from the expanded MatchMaker item pool.

In interpreting correlation coefficients we will use Cohen's (1988) guidelines. Correlations of approximately .10 are considered small, correlations of approximately. 30 are considered medium and correlations of approximately .50 are considered large. Generally speaking, test-retest reliability estimates are expected to be larger the shorter the latency between initial and repeated measurement.

Results
Test-Retest Reliability

We conducted four sets of analyses to examine the temporal stability or test-retest reliability of individual MatchMaker items as well as the Holland-like scales derived from the previous analysis. The four sets of tables below outline our findings and connect these outcomes to extant literature on the stability of vocational interests and the test-retest reliability measurement of these constructs.

Core Items

Our first analysis examined the test-retest reliability of the 39 core items of the MatchMaker instrument. We calculated the test-retest reliability coefficient for each of the 39 core items across the four samples. Below we report the average test-retest reliability for all 39 items, as well as the specific

items that demonstrated the highest and lowest correlation coefficients. Mean values are reported in this table with the sample sizes for these analyses noted in the last column.

The item level analysis of test-retest reliability demonstrates large correlations for all items. As expected, we found higher correlations over a shorter time frame (1-3 weeks) compared to the longer test-retest period in our samples (3-4 months).

	Mean	Maximum	Item #	Minimum	Item #	Sample
HS 7-21	0.66	0.75	1	0.54	7	36820
HS 90-120	0.59	0.70	1	0.43	7	12150
College 7-21	0.68	0.76	1	0.55	7	1047
College 90-120	0.61	0.73	16	0.42	20	182

Item #	Item Content
1	Working with children
7	Doing hands-on work not needing new skills
16	Providing medical care and treatment to people
20	Obeying orders and instructions

Core Items by RIASEC

We specifically examined the test-retest reliability of the 39 core MatchMaker items when organized according to a RIASEC Holland structure (described in detail in the MatchMaker Scale Evaluation Report). To accomplish this analysis we created 6 RIASEC scales using the identified factor structure of the 39 core items. We then calculated the test-retest reliability of these scales with the results presented below.

Examination of the core items by the RIASEC scales finds correlations between .56 and .79 among the four data sets. Predictably, shorter test-retest latencies generally resulted in larger coefficients. Overall these correlations compare favorably with other instruments' measurement of interests over time. Zarella and Schurerger (1990) reported an average test-retest reliability correlation of .67 for seven different instruments across 83 different samples.

	R	I	A	S	E	C	Sample
HS 7-21	0.79	0.75	0.75	0.77	0.73	0.71	36820
HS 90-120	0.75	0.69	0.69	0.74	0.67	0.65	12150
College 7-21	0.78	0.74	0.74	0.77	0.76	0.77	1047
College 90-120	0.77	0.56	0.74	0.72	0.78	0.65	182

Expanded Items

The next analysis explored the test-retest reliability of MatchMaker items using the newly identified factor structure that considered all 116 items to create RIASEC scales. Similar to the first analysis we

started with examination of test-retest reliability at the item level followed by an analysis of coefficients using pre-established factors. The RIASEC factor structure used in this analysis included seven items in each of the scales. The table below notes the average test-retest correlations for the seven items in each scale. As in the first table in this section, these are the Mean correlations of included items.

The results of this analysis are smaller average correlation coefficients for the expanded RIASEC scales as compared to the core 39 items. While large by Cohen's definition these items have somewhat less stability over time. It is important to note that each of the samples is less than half the size of the samples for the core items.

	R	I	A	S	E	C	Sample
HS 7-21	0.55	0.55	0.55	0.50	0.55	0.54	17293
HS 90-120	0.44	0.45	0.44	0.37	0.47	0.45	5687
College 7-21	0.48	0.53	0.56	0.43	0.56	0.51	510
College 90-120	0.40	0.45	0.49	0.30	0.46	0.41	97

Expanded Items by RIASEC

Our final examination of test-retest reliability examined the stability of the new RIASEC scales over time. Similar to the average reliabilities at the item level, correlations here are slightly smaller compared to core item analyses. The two samples measured over a 7-21 day time span do demonstrate respectable reliability with all correlations in the .60s to low .70s.

	R	I	A	S	E	C	Sample
HS 7-21	0.69	0.71	0.71	0.63	0.71	0.67	17293
HS 90-120	0.61	0.62	0.63	0.53	0.63	0.60	5687
College 7-21	0.65	0.69	0.67	0.62	0.69	0.65	510
College 90-120	0.59	0.67	0.56	0.49	0.65	0.46	97

Test-Retest Reliability Summary

Our analysis of MatchMaker finds that the instrument reliably measures career interests over two periods of time. The RIASEC structure of core items and to a slightly lesser extent, expanded items, compare favorably to other vocational interest instruments (Zarella & Schuerger, 1990). The items and scales drawn from the 39 core MatchMaker items demonstrate slightly more stable measurement of career interests as compared to the expanded item set. This finding may be impacted by the differences in sample size or it may reflect other response tendencies such as fatigue experienced during the completion of extended item sets.

To further understand this phenomenon, we ran a post-hoc analysis of the average test-retest reliability coefficients of sets of items that derive from the (a) core item set and (b) expanded item set *and* contribute to the newly formed factors. A table displaying these results is presented below. Parentheses

next to scale designations reflect the number of items in the core/expanded set contributing to each scale.

	R (2/5)	I (2/5)	A (3/4)	S (1/6)	E (3/3)	C (2/5)
HS 7-21 Core	0.71	0.69	0.67	0.69	0.62	0.68
Expanded	0.48	0.50	0.46	0.47	0.47	0.49
HS 90-120 Core	0.65	0.63	0.60	0.63	0.55	0.62
Expanded	0.36	0.37	0.33	0.32	0.35	0.38
College 7-21 Core	0.70	0.72	0.68	0.69	0.68	0.70
Expanded	0.40	0.45	0.46	0.39	0.39	0.44
College 90-120 Core	0.65	0.59	0.67	0.61	0.62	0.66
Expanded	0.30	0.40	0.35	0.25	0.24	0.31

Clearly items from the core have considerably higher average test-retest reliability coefficients compared to those deriving from the expanded item set and contributing to these factors.

These findings suggest possible limitations to the use of expanded items for any purpose, including that of developing Holland-like factors or scales given their relative lack of temporal stability relative to items in the core.

Methodology
Construct Validity

In order to establish the construct validity of the MatchMaker instrument is was necessary to gather responses on both the MatchMaker core and expanded item set and another measure of career interests in a single sample of students. To accomplish this, CECS established an online survey that contained by MatchMaker items and the full item set from the O*NET Interest Profiler, a commonly used interest inventory available in several competitor products that provides results using 6 Holland scale scores. Adult college samples were obtained through participation in the University of Utah Educational Psychology research participant pool. College students taking introductory study skills and psychology courses were solicited for participation. High school samples were obtained through collaboration with CareerCruising. Existing high school clients were asked to refer their students to the online survey. College students were incented through the administration of course-based extra credit whereas high school students were incented through the provision of $10.00 Amazon.com gift cards. Sample size limitations prohibit separate construct validity analysis of high school and college samples. As such, the samples were combined together and resulted in 523 usable records.

Two analyses were conducted. In the first set of analyses Holland factors were created using only the 39 core MatchMaker item set as described in previous report. In the second set of analyses, Holland factors were created using the expanded MatchMaker item set as previously described. Resulting MatchMaker Holland scale scores from both item sets were then compared to O*NET Interest Profiler scale scores.

Results
Construct Validity

The multi-trait multi-method matrix showing correlations among the scale scores from the two instruments using only core MatchMaker items is shown below.

Note that the two instruments are scaled in opposite directions meaning that strong negative correlations actually reflect scale correspondence. Moderate correlations (e.g., 0.33 to 0.58) can be
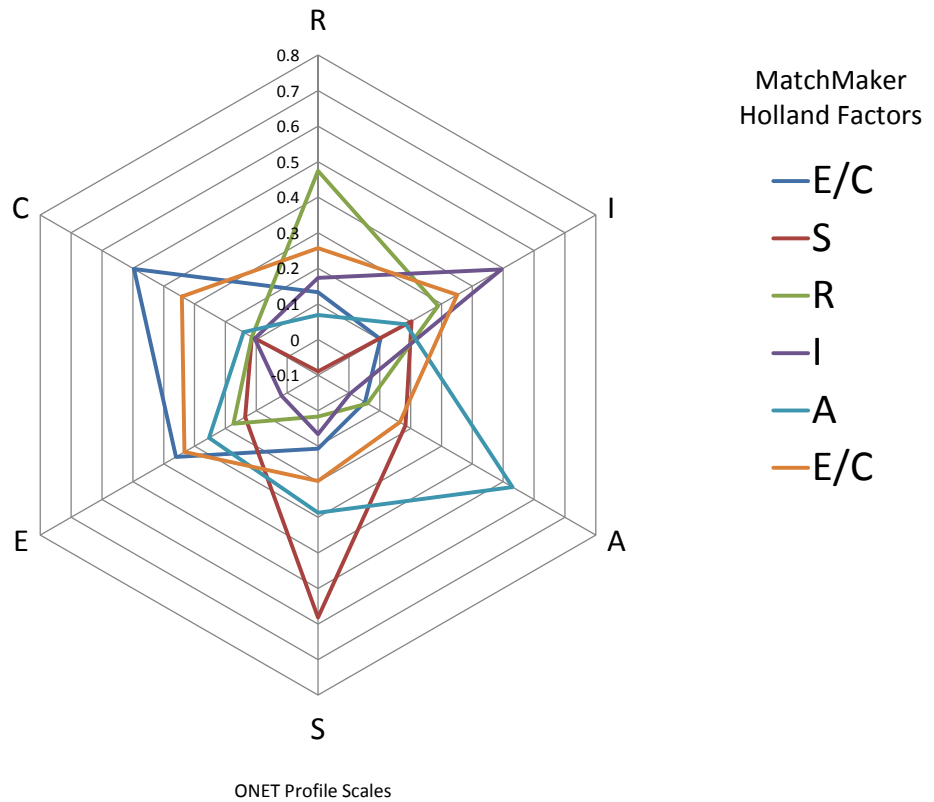
Correlations

		ONET_R	ONET_I	ONET_A	ONET_S	ONET_E	ONET_C
CC_E/C	Pearson Correlation	-.133[**]	-.102[*]	-.052	-.107[*]	-.359[**]	-.496[**]
	Sig. (2-tailed)	.003	.023	.248	.017	.000	.000
CC_S	Pearson Correlation	.090[*]	-.203[**]	-.183[**]	-.582[**]	-.136[**]	-.114[*]
	Sig. (2-tailed)	.046	.000	.000	.000	.002	.011
CC_R	Pearson Correlation	-.474[**]	-.289[**]	-.060	-.017	-.174[**]	-.119[**]
	Sig. (2-tailed)	.000	.000	.182	.702	.000	.008
CC_I	Pearson Correlation	-.173[**]	-.494[**]	.004	-.066	-.018	-.105[*]
	Sig. (2-tailed)	.000	.000	.923	.145	.691	.019
CC_A	Pearson Correlation	-.069	-.185[**]	-.530[**]	-.287[**]	-.253[**]	-.141[**]
	Sig. (2-tailed)	.127	.000	.000	.000	.000	.002
CC_E/C	Pearson Correlation	-.257[**]	-.351[**]	-.165[**]	-.198[**]	-.333[**]	-.341[**]
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

observed between like-scales on the two instruments. In contrast, considerably smaller correlations are typically observed between non-like scales. A notable exception to these observations are the fact that the first and last MatchMaker scales are relatively poorly defined and appear to capture item content associated with both Enterprising and Conventional scales on the O*NET Interest Profiler. In contrast, the R, I, A, and S scales on the MatchMaker instrument appear to be very clearly defined measure of these underlying interest construct.

The figure below represents these data in another way. For the purposes of this figure, all correlations have been transposed to the opposite sign (negative correlations are displayed as positive and vice versa) to clarify the relations observed among scales on the two instruments.

Correlations Between MatchMaker Holland Factors and ONET Interest Profile Scale Scores

MatchMaker Holland Factors
— E/C
— S
— R
— I
— A
— E/C

ONET Profile Scales

The axis of this figure displays O*NET Interest Profiler scales and the radar plots show the relations of MatchMaker scale scores to O*NET Interest Profiler scales. A clear pattern of construct validity emerges as one examines this plot. In all cases, like-scales clearly emerge as having the highest overall relationships while non-like-scales display considerably lower overall relationships. Results from this analysis can be compared to previous efforts at examining construct validity of interests using two distinct measures (Sullivan & Hansen, 2004). Correlations between like-scales (Strong Interest Inventory and Campbell Interest and Skills-Survey) in that study ranged between .45 and .75.

The multi-trait multi-method matrix showing correlations among the scale scores from the two instruments using the expanded core MatchMaker items is shown below.

Correlations

		ONET_R	ONET_I	ONET_A	ONET_S	ONET_E	ONET_C
CC_S	Pearson Correlation	.047	-.209[**]	-.223[**]	-.668[**]	-.153[**]	-.138[**]
	Sig. (2-tailed)	.298	.000	.000	.000	.001	.002
CC_R	Pearson Correlation	-.565[**]	-.311[**]	-.078	.027	-.225[**]	-.171[**]
	Sig. (2-tailed)	.000	.000	.086	.557	.000	.000
CC_A	Pearson Correlation	-.070	-.213[**]	-.557[**]	-.285[**]	-.215[**]	-.098[*]
	Sig. (2-tailed)	.123	.000	.000	.000	.000	.030
CC_I	Pearson Correlation	-.220[**]	-.627[**]	-.154[**]	-.178[**]	-.095[*]	-.094[*]
	Sig. (2-tailed)	.000	.000	.001	.000	.035	.038
CC_C	Pearson Correlation	-.122[**]	-.157[**]	-.090[*]	-.051	-.223[**]	-.421[**]
	Sig. (2-tailed)	.007	.000	.045	.260	.000	.000
CC_E	Pearson Correlation	-.147[**]	-.181[**]	-.116[**]	-.214[**]	-.389[**]	-.397[**]
	Sig. (2-tailed)	.001	.000	.010	.000	.000	.000

Moderate to strong correlations (e.g., 0.39 to 0.68) can be observed between like-scales on the two instruments. In contrast, smaller correlations are typically observed between non-like scales. Although the MatchMaker conventional scale more clearly delineates itself in this analysis the MatchMaker enterprising scale continues to be relatively undifferentiated sharing variance with both the O*NET Interest Profile enterprising and conventional scales. The MatchMaker R, I, A, S, and C scales all show considerably higher evidence of construct validity when using the expanded set of items and appear to clearly defined these underlying interest constructs.

The figure below represents these data in another way. For the purposes of this figure, all correlations have been transposed to the opposite sign (negative correlations are displayed as positive and vice versa) to clarify the relations observed among scales on the two instruments.

As with the figure above, the axis of this figure displays O*NET Interest Profiler scales and the radar plots show the relations of MatchMaker scale scores to O*NET Interest Profiler scales. A clear pattern of construct validity emerges as one examines this plot. In all cases, like-scales clearly emerge as having the highest overall relationships while non-like-scales display considerably lower overall relationships. The increased overall magnitude of like-scale intercorrelations is clear comparing this to the previous figure. Moreover, results from this analysis begin to approach construct validity coefficients previously reported by Sullivan & Hansen (2004).

Correlations Between MatchMaker Holland Factors and ONET Interest Profile Scale Scores
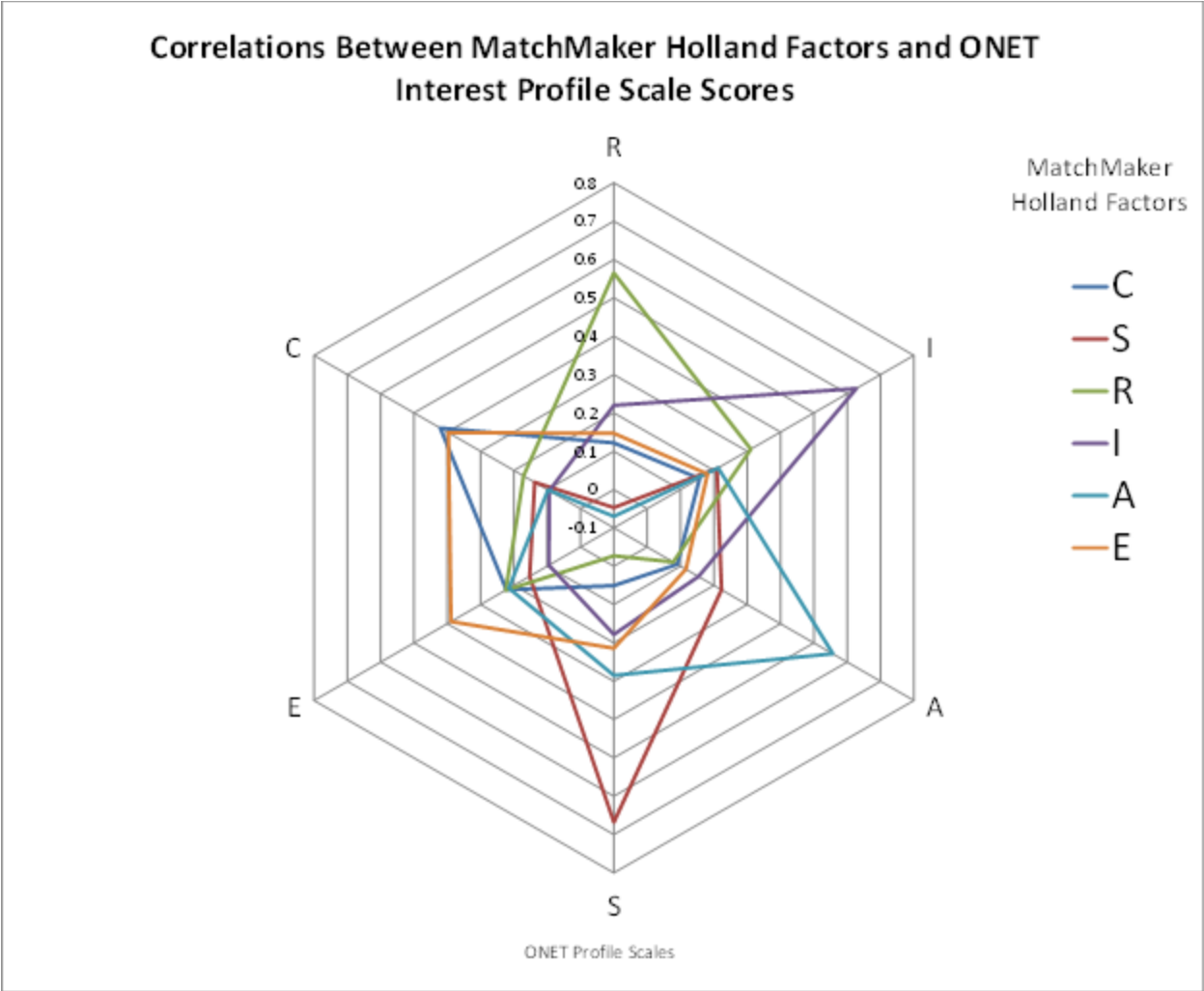
Construct Validity Summary

Results from the examination of the construct validity of response to the MatchMaker inventory suggest that the underlying Holland-like factors are a viable method of presenting results of examinees responses. A clearly differentiated pattern of correlations can be observed in the tables and figures above with Like-scale intercorrelations ranging from 0.33 to 0.68 and non-like-scale correlations in almost all cases being slightly or significantly lower. It should be noted that notably stronger evidence exists for factors/scales created using the expanded item sets. These conclusions can be supported by two observations. First, the MatchMaker conventional scale is more clearly differentiated using the expanded item set and second, the overall magnitude of the like-scale correlations is higher.

The fact that these results argue for the use of the expanded item set stands in stark contrast to results from the reliability analysis described above where use of expanded items jeopardized the test-retest reliability of factors/scales.

The practical implications of these findings taken together are that increased confidence can be placed in Holland-like scales when using the expanded item pool but scores on those scales will generally be less stable (albeit still acceptable) over time than would scales created using only the core item set.

One caveat to these conclusions is critical to consider. If CareerCruising decides to replace existing poorly performing core items with expanded set items in an effort to create a new core that best describes Holland RIASEC constructs, the possible influences seen on the instability of expanded set items (e.g., fatigue) may be mitigated.

General Conclusions

The existing MatchMaker item set displays adequate to strong temporal stability (test-retest reliability) and in patterns that would be expected psychometrically. When factors are examined for their stability over time, factors developed using only core items outperform those produced using the core plus expanded item set. Caution should be exercised using the expanded items in their current placement and configuration. Doing so will result in lower test-retest reliability estimates (and inferences).

Preliminary but strong evidence exists in support of the construct validity of response to items on the MatchMaker instrument. We conclude that Holland-like scales on the MatchMaker instrument are possible and defensible based on our findings in this and previous report. Considerably stronger evidence exists for the use of the expanded item set in developing such scales should limitations with the temporal stability of those items be resolved through repositioning of expanded set items.

References

Sullivan, B. A., & Hansen, J. C. (2004). Evidence of construct validity of the interest scales on the Campbell Interest and Skill Survey. *Journal of Vocational Behavior, 65,* 179-202.

Zarella, K. L., & Schuerger, J. M. (1990). Temporal stability of occupational interest inventories. *Psychological Reports, 66,* 1067-1074.

Career & Educational Consulting Services, LLC (CECS) provides evaluation, training, and consultation services to a broad range of units and programs in secondary and post-secondary institutions, scholarship foundations, and career guidance companies.

Paul Gore, Ph.D., has 16 years of experience promoting the academic and career success of high school and college students. He consults with high schools, colleges, and universities in the U.S. and abroad on the development and implementation of student academic and career success programs. Dr. Gore has conducted extensive research on the factors that predict college student academic and career success and he has authored over 50 book chapters and journal articles and delivered over 100 professional presentations and invited addresses. He holds a faculty position in the Department of Educational Psychology at the University of Utah where he also serves as the Director of Institutional Research and Student Success Special Projects Coordinator. Prior to moving to Utah, Dr. Gore held a position as Director in the Research Division of ACT, Inc, a leading college admissions testing company. He is actively involved in the training and supervision of professional high school counselors and counseling psychologists. Dr. Gore holds a Ph.D. in Counseling Psychology from Loyola University of Chicago.

Wade Leuwerke, Ph.D., earned a Ph.D. in Counseling Psychology from Southern Illinois University at Carbondale. He has 11 years of experience advancing the academic and career success of students and individuals. Dr. Leuwerke consults with colleges and universities on the use of data driven strategies to increase academic performance and student retention. He has experience working with students in high school and college, as well as adults, to develop strategies for academic and career success. Dr. Leuwerke has over 30 journal articles, book chapters and presentations. His research has focused on examining college retention, career development, computer assisted career guidance systems, and the role of technology in career development processes. Dr. Leuwerke is currently an associate professor in the Counselor Education program at Drake University, Des Moines, Iowa. He has experience as a project manager and research consultant to ACT, Inc., and Kuder, Inc.