APPENDIX A

Standardized Measures Analysis Report

To answer research questions concerning the degree of self-change the participants reported or demonstrated, we collected a range of demographic data and also scored the Subject-Object Interview (SOI) and three standardized measures of psychological status. Participants' scores on these standardized measures—the Satisfaction with Life scale (SWL), the Personal Efficacy Beliefs Scale (PEBS), and the Locus of Control scale (LOC)—were specifically intended to help us find and describe changes in participants' overall satisfaction and confidence as they engaged in the programs we studied, using widely accepted measures of these variables. Using a variety of appropriate statistical techniques (primarily simple and multiple regression analyses), we analyzed the numerical, demographic, and psychological data we collected, looking for statistically significant differences in the populations at our three sites and correlational relationships among the demographic, paper and pencil, and developmental variables. This section presents our methods and our findings concerning these quantitative analyses.

The Variables

Research on adult development persistently reports a positive association between participants' years of education and the demonstration of higher-stage reasoning. A variety of studies have also looked at the relationship between gender and adult development, with contradictory reports (using a variety of methods) linking either gender to higher-stage cognitive or moral reasoning. Also, several studies have suggested provocative relationships between SES status overall and level of development across several domains of adult life (e.g., work, parenting, interpersonal relationships). To support our exploration of the important relationships among key demographic variables and developmental stage and also to pursue questions on the relationship of other variables to life satisfaction and locus of control, at our initial visit to each site, we gathered a variety of demographic information about our participants. These include Age at First Visit, Gender, Marital Status, Number of Children, ESOL Status, Years in the United States, Years of Own Education, Years of Mother's Education, and Years of Father's Education.

Because we were unsure which relationships among parents' education and success might matter most, we also created and tested variables for **Years of Best- Educated Parent's Education** and **Years of Same-Gender Parent's Education**. Because we thought there might be relationships among variables that depended more on whether or not a participant *was* a parent than incrementally on number of children, we also created a dichotomous variable for **Parental Status**.

In addition to these purely demographic variables, we also attended to variables related to the program that participants were in, specifically the **Site** itself, and **Months Already in the Program** at first visit. Finally, we administered the Subject-Object Interview (**SOI**) (Lahey, Souvaine, Kegan, Goodman & Felix, 1988), and the three paper and pencil measures—**PEBS**, **SWL**, and **LOC**—at both our initial and final visits and created derived variables for changes in scores for these variables. These derived variables represent simply the differences between time one and time last scores on each measure.

As our analysis would use regression tools that relied on more or less normal, linear distributions of variables, we examined these distributions for all of our variables. We found all but two of the variables to be normally distributed. These two—**Years in the United States**, and **Number of Children**—were positively skewed. To adjust for the unreliability of our measures of central tendency and to address potential problems with our regression analyses, we transformed them using a logarithmic transformation adjusted to avoid undefined values (LOG(1+*RawValue*)), conducted our analyses with these transformed variables, and then untransformed them ((10^*TransformedValue*)–1) to report our results.

Table 1 reports means, standard deviations, medians, and number of participants contributing to these (N) for each of the demographic variables, both within each site and overall. Notice that the reported standard deviations for our transformed variables are substantially lower than those for the raw variables, indicating that this procedure served to linearize these variables.

Table 1: Descriptive Statistics for Demographic Variables

Variable	Polaroid Mean	σ	Median	N	Evenstart Mean	σ	Median	N	BHCC Mean	σ	Median	N	Overall Mean	σ	Median	N
Site				18				20				17				55
Age	41.94	7.58	44.00	17	32.67	6.14	33.00	18	25.15	6.62	24.00	13	33.92	9.46	33.00	48
% Female	0.50	0.51	0.50	18	0.70	0.47	1.00	20	0.59	0.51	1.00	17	0.60	0.49	1.00	55
Marital status	.12 single .65 married .24 divorced			19	.055 single .90 married .055 divorced				.86 single .14 married				.3 single .6 married .1 divorced			
Number of Children	2.59	1.66	3.00	17	2.95	1.90	3.00		0.00	0.00	0.00	14	2.00	1.96	2.00	50
# Children (transfrmd)	1.96	.85	2.46	17	2.38	.66	2.46	19	0	0	0	14	1.22	1.03	1.00	50
% Parents	0.88	0.33	1.00	17	1.00	0.00	1.00	19	0.00	0.00	0.00	14	0.68	0.47	1.00	50
Years in the U.S.	21.12	12.09	18.00	17	10.11	7.25	9.00	19	3.14	3.90	2.00	14	11.90	11.12	10.50	50
Years in the U.S. (transfrmd)	15.13	1.48	17.50	17	6.64	1.62	9.00	19	1.70	1.04	2.00	14	6.07	2.07	9.00	50
Years Own Ed	9.47	3.74	10.00	17	9.53	3.59	10.00	17	12.21	2.15	12.00	14	10.29	3.46	11.00	48
% ESOL	0.76	0.44	1.00	17	0.82	0.40	1.00	11	1.00	0.00	1.00	14	0.86	0.35	1.00	42
Mos Pgm	NA	NA	NA	0	19.74	17.42	13.00	19	9.29	7.56	7.50	14	15.30	14.88	10.00	33
Mother Ed	4.19	3.75	5.00	16	3.71	6.16	0.00	7	9.00	5.96	8.00	13	5.83	5.53	5.00	36
Father Ed	7.50	3.98	5.00	16	5.86	5.46	5.00	7	10.83	5.36	12.00	12	8.31	5.03	7.00	35
Most Parent's Ed	7.63	4.00	5.00	16	7.13	6.20	6.00	8	11.00	5.48	12.00	13	8.70	5.21	8.00	37
Same Gender Parent Ed	6.19	4.29	5.00	16	5.50	5.55	6.00	8	9.42	6.43	8.50	12	7.11	5.47	5.50	36

As reported in the main body of the monograph, clear differences in mean age across sites are evident, with participants in the Polaroid site representing the oldest group on average and the learners at the BHCC site the youngest. No participants at BHCC report having children.

Reliability Analysis

In administering the paper and pencil measures with a population for whom English is not generally their primary language, we found ourselves unsure whether participants accurately and consistently understood the questions that make up these measures. Thus, we had doubts about whether the measures would be sufficiently reliable—whether participants' answers would be sufficiently robust—to be used at all. Therefore, we began with an analysis of the reliability of the measures themselves with this population before undertaking the more substantive correlational and longitudinal analyses.

The Personal Efficacy Beliefs Scale (PEBS) is a 10-item measure of self-efficacy (Riggs, Warka, Babasa, Betancourt & Hooker, 1994); the Locus of Control (LOC) scale is a seven-item measure of locus of control; and the Satisfaction With Life scale (SWL) is a five-item measure of general life satisfaction (Diener, Emmons, Larsen, & Griffin, 1985; Pavot, Diener, Colvin, & Sandvik, 1991). These measures were administered at the first and last data collection visits at each of the three sites. At Even Start for both visits, and at BHCC for the final visit, we also extended the SWL by doubling and modifying each of the original questions to address satisfaction with the primary role of interest (parent, student), yielding a 10-item scale that really consisted of two parallel subscales, SWLife and SWRole, though the SWRole scale proved unsatisfactory, as described below.

In the published versions of each of these scales, responses are to be on a seven-point Likert type scale, but we found after administering these at the Polaroid site that this complex scale seemed too difficult to understand for the limited English proficient participants, so we changed the response format to a five-point Likert type scale at Even Start and BHCC for both data collection visits. In addition, we included graphics on the scale to help participants understand the meaning of the ratings. Having used a seven-point format at Polaroid at the initial visit, we kept it for the final visit.

Method

The reliability analysis for these scales consisted of several steps:

1) We entered and then cleaned the data by data re-orienting responses when questions were phrased in the negative[1]; converting responses on the seven-point scale at Polaroid to a five-point format so that data across all three sites could be used in the same analysis[2]; separating out the SWLife and SWRole scales; and dealing with missing values or values participants entered mistakenly.

2) We calculated the Cronbach α statistic for each scale at each administration. This measures the correlation between participants' responses on each item and the total of the other items on that scale. It is a measure of internal consistency of the measure in that each item on the scale is supposed to contribute positively to the overall score on that measure so they should all be positively and highly correlated. We compared these statistics both with the published values for each measure (all > .8) and with general guidelines for what counts as a reasonable value for these statistics ($\geq$ .6).

3) We also examined whether the Cronbach α statistic would increase or decrease upon the deletion of each item in the measure—again, if excluding an item would increase the consistency of the remaining items, one could argue it should be excluded to make the measure more robust. In our case, this argument needed to be balanced against the desire to remain true to the original scales (see below for results of this procedure).

4) Finally, we created composite scores by summing all the included items and dividing by the number of items. This method standardizes the scores to be consistent with the original 1 to 5 Likert scale so scores can be compared across sites and administrations.

Results and item deletion analysis

The following chart was used to evaluate the internal reliability of the measures using the Cronbach α statistic. It also lists how the a statistic would change upon deletion of the least correlated item and notes that item.

[1] PEBS questions 1, 5, 7, & 9; LOC questions 1, 2, 3, 5, & 7.
[2] This yields seven possible responses distributed across the five-point range, viz. at 1, 1.66, 2.33, 3, 3.66, 4.33, and 5.

Time	Scale	Cronbach α	N	Proposed deletion	α on deletion
Initial	SWLife	.510	50	8	.584
Final	SWLife	.630	31	10	.761
Initial	SWRole	-.300	11		
Final	SWRole	.353	18	9	.667
Initial	LOC	.352	50	2	.450
Final	LOC	.600	37	6	.625
Initial	PEBS	.654	51	7	.669
Final	PEBS	.728	31	9	.747

Using this chart, we made decisions about item deletion, considering and balancing: 1) Could the measure as it stood be sufficiently reliable to use even if its reliability could be increased by item deletion? If so, it seemed advantageous to leave the measures as published. 2) How much of an increase in reliability would item deletion yield? If the increase was small, again it seemed better to leave the measure as published rather than modifying it.

For criterion #1, we decided that an α value greater than or equal to .6 was sufficiently robust not to alter the measure. The PEBS at both administrations, the LOC at the final administration, and the SWLife scale at the final administration fit this criterion, so we let them stand with all of their items included in the composite score. It could be argued that the dramatic increase in α (an increase of .13) upon the elimination of item 10 in the final SWLife makes such an elimination a reasonable choice; however, we chose not to do so.

Difficulties with our initial administration of our constructed SWRole scale raised our concerns about its reliability. We made the conservative decision to eliminate it as a measure for the study. Although the final administration of this measure could be made reliable by deleting item 9, we decided to drop this measure as well, both because it was not administered across all three sites and because of the lack of a longitudinal comparison score.

Deleting item 8 from the initial administration of the SWLife measure produces a substantial (.074) increase in α and a modified measure that is sufficiently reliable to use. Although the reliability estimate, at .584, is not quite up to our .60 criterion, no further item deletions would increase the internal reliability at all, so we settled on this measure with just item 8 deleted.

Deleting item 2 from the initial administration of the LOC measure also yields a substantial (.118) increase in α, but the modified measure is not yet sufficiently reliable to use (α = .45). However, no further item deletions would increase the internal reliability at all so we had to settle with deleting item 2 and a much less than ideal internal reliability of .45. However, the relative unreliability of this measure at this administration must be considered as we examine the meaning of our correlational results.

Summary of reliability analysis

We conducted a Cronbach α test of reliability on the paper and pencil measures used in this study and found that though these estimates are lower than the published values, the measures are sufficiently reliable to be used as is in all of their final administrations and for the PEBS, also in its initial administration.

By deleting uncorrelated items from the SWLife scale in its initial administration, this scale can also be made sufficiently reliable to use in further analyses. A similar effort with the initial administration of the LOC scale leaves more ambiguous results. The SWRole scale, constructed for this project, cannot be so adjusted in its initial administration, and therefore, we dropped it from the study.

Although we will examine differences in these scores across administrations and sites, notice that the Cronbach α reliability values consistently go up across administrations of the same scale. This may serve as one kind of evidence of the increasing linguistic competence of the participants as they were increasingly able as a group to understand the language underlying these questions and thus, better able to respond consistently and reliably.

Distribution of Psychological Variables

Having checked on the reliability of these measures, we examined their distribution before turning to relationships among these many variables. Table 2 shows the mean, standard deviation, median, and associated N for the three paper and pencil measures—LOC, PEB, and SWL—and the Subject-Object Interview (SOI) scores. For each measure, statistics are listed for both the initial and final data collection visits, as well as for a derived change in score variable computed by subtracting initial from final score. We also checked the distributional characteristics of these variables and found them all to be normally, linearly distributed.

Table 2: Descriptive Statistics for Psychological Variables

	Polaroid				Evenstart				BHCC				Overall			
Variable	**Mean**	**σ**	**Median**	**N**	**Mean**	**σ**	**Median**	**N**	**Mean**	**σ**	**Median**	**N**	**Mean**	**σ**	**Median**	**N**
SOI Initial	2.73	0.55	2.50	15	2.94	0.47	3.00	19	3.08	0.33	3.00	16	2.92	0.47	3.00	50
SOI Final	2.86	0.54	2.60	17	2.99	0.40	3.00	14	3.03	0.42	3.00	10	2.95	0.46	3.00	41
SOI Chng	0.13	0.24	0.10	15	0.01	0.09	0.00	14	-0.07	0.19	0.00	9	0.04	0.20	0.00	38
LOC Initial	3.26	0.53	3.11	16	3.62	0.66	3.67	17	3.58	0.66	3.50	17	3.49	0.63	3.50	50
LOC Final	3.65	0.78	3.48	17	3.81	0.66	3.86	12	3.64	0.50	3.50	8	3.70	0.68	3.57	37
LOC Chng	0.32	0.93	0.11	15	0.20	1.08	0.13	12	-0.15	0.36	-0.10	8	0.17	0.89	0.00	35
PEB Initial	3.81	0.72	3.73	17	3.92	0.38	3.80	17	3.78	0.53	3.70	17	3.84	0.55	3.73	51
PEB Final	4.33	0.59	4.47	17	4.10	0.55	4.00	12	3.91	0.57	3.75	8	4.16	0.58	4.00	37
PEB Chng	0.54	0.53	0.33	16	0.12	0.75	0.00	12	-0.15	0.60	-0.15	8	0.27	0.69	0.14	36
SWL Initial	3.45	0.87	3.53	16	3.41	0.67	3.60	17	2.93	0.66	3.20	17	3.26	0.68	3.23	50
SWL Final	3.77	0.97	3.93	17	3.73	0.54	3.80	12	3.38	0.45	3.30	8	3.70	0.75	3.80	36
SWL Chng	0.20	1.15	0.40	15	0.38	1.03	0.18	12	0.33	0.72	0.30	8	0.31	0.95	0.47	34

Site Specific Differences

Having examined the distributions of these variables both overall and at each of the sites, we then explored whether differences among sites were statistically significant. We found differences among the sites in several of the demographic and psychological variables—some at the level traditionally considered statistically significant (pΔ.05—listed in **bold**) and some bordering on this level (.05<pΔ.10). In Table 3, we list site specific averages, the value of the F test and its degrees of freedom, the associated p value, and the value of R^2 (it ranges from 0 to 1). Thus, 1/2 the variation in Age can be predicted merely by knowing a student's site.

Table 3: Site Specific Differences in Demographic Variables

Variable	Polaroid average	Evenstart average	BHCC average	F	df	P value	R^2
Age	41.9	32.6	25.1	**22.89**	**2, 46**	**.0001**	**.50**
Years in the U.S. (direct average)	21.1	10.1	3.1	**17.41**	**2, 48**	**.0001**	**.43**
Years in U.S. (transformed)	18.0	7.5	2.3	**25.24**	**2, 48**	**.0001**	**.50**
Own Education in years	9.5	9.5	12.2	**3.35**	**2, 48**	**.044**	**.13**
Mother's Education in years	4.2	3.7	9.0	**3.92**	**2, 36**	**.030**	**.19**
Father's Education in years	7.5	5.9	10.8	2.82	2, 33	.075	.15
# Kids (direct average)	2.6	2.9	0	**17.01**	**2, 48**	**.0001**	**.42**
# Kids (transformed)	2.2	2.6	0	**13.89**	**2, 48**	**.0001**	**.64**
Parent status	88%	100%	0%	**121.39**	**2, 48**	**.0001**	**.73**

Thus, as noted above, students at BHCC are, on average, younger and more recently arrived in the United States, have no children, and have more years of education themselves as do both their parents. This differs significantly from both the other sites. At Polaroid as well, students are, on average, older, have been in the United States longer, and have slightly better educated fathers than students at Even Start. There are no statistically significant differences by site in Gender or ESOL status.[3] Because marital status has three possible values at each of the three sites, we conducted a chi-square test to determine that marital status does, indeed, differ significantly by site (χ^2=32.9, df=4, p<.0001). Again, the difference here is primarily that BHCC students are generally single, and that is rare for students at the other sites.

There are also a few site-specific differences in the psychological variables we measured, or their derivatives. These are displayed in Table 4, below.

Table 4: Site Specific Differences in Psychological Variables

Variable	Polaroid average	Evenstart average	BHCC average	F	df	P value	R^2
Change in Efficacy score	.50	-.02	-.15	2.82	2, 34	.070	.20
Change in SOI score	.13	.01	-.07	**3.50**	**2, 35**	**.041**	.17
ΔSOI w/o outlying POL stdt	.09	.01	-.07	2.97	2, 34	.065	.15

Thus, students at Polaroid had, on average, higher increases in both their PEBS scores and in their SOI score over the time we studied them. Examining these results, it seemed that increases in SOI score might be attributable to a single Polaroid student's substantial increase in SOI score from initial to final time. However, removing this student from the data set (last line in Table 4) we still find a strong and nearly statistically significant relationship between Δ SOI and site.

Though these differences are small, they *are* statistically significant or nearly so and it is worth asking what characteristics of the Polaroid program over the time period studied in comparison with the other sites led to these changes in efficacy and constructive-developmental level? Was it just

[3] For the purposes of our analysis, ESOL status is defined as a binary variable.

that we studied Polaroid students over a longer time period than the other sites? Or did something about the program itself that promote both Efficacy Beliefs and the development of cognitive complexity?

Demographic Predictors of Psychological Variables

By fitting simple and multiple regression models, we explored what demographic variables might predict SOI or other paper and pencil measures. Many that we would expect to do not, even controlling for differences by site. Thus, Years of Own Education, Age, and Years in the United States fail to predict initial SOI scores, nor does Gender or Marital status. Most years of parents' education does not predict SOI, nor does Years of Father's Education, though Years of Mother's Education does with coefficients that are borderline in their statistical significance ($p=.065$, $R^2=.112$).

$$\textit{Initial SOI} = 2.73 + .032\ \textit{Mother's Education}$$

This means that each additional year of a student's Mother's Education is associated with an SOI score that is .032 higher. The relationship between SOI and Mother's Education (and not Father's Education or Same-gender parent's education or Most parents' education) is interesting. Does mother's education still mean something about the value of education in a family? And how would that be associated with increased SOI scores?

In our data set, parent status also predicts initial SOI, with parents, on average, seeming to have less cognitive complexity than non-parents. This, however, turns out to be an anomaly of the distributional characteristics of our data (all parents at Even Start, no parents at BHCC, and just a few non-parents with higher SOI scores at Polaroid), so we've ruled out any generalization to a larger population for this non-intuitive result. There are no other statistically significant relationships between demographic variables and any of the psychological variables.

Psychological Variables Predicting SOI

We began by examining how the paper and pencil measures might be correlated with SOI scores. Again, we fitted several simple and multiple regression models, looking for those with statistically significant values both for the overall equation (F statistic) and for each of the individual regression coefficients. When a variable is statistically significant in a multiple regression model, it means that it contributes to the predictive power of the model even after controlling for the other variables in the model.

Table 5 lists regression coefficients for the several statistically significant or borderline significant models relating initial scores on psychological variables and site to initial SOI score. The coefficients from the table can be interpreted, when present, as filling in for coefficients in equations of the form:

$$InitialSOI = \text{Intercept} + A \times InitialSWL + B \times InitialLOC + C \times BHCC + D \times EVST$$

There are two dichotomous site variables, always entered together—BHCC indicating a student is at BHCC, and EVST indicating that a student is at Even Start. The Polaroid site does not have a separate variable because it is described by those who are neither BHCC nor EVST.

Table 5: Predicting Initial SOI by Other Psychological Variables and Site

		A	B	C	D				
Model	**Intercept**	**SWL Initial**	**LOC Initial**	**BHCC**	**EVST**	**F statistic**	**p**	**Df**	**R^2**
I	3.42	–.147 ~				3.22	.080	1, 43	.070
II	3.08	–.113		.317 ~	.307 ~	2.66	.063	3, 41	.161
III	2.17		.220 *			**4.28**	**.045**	1, 41	.095
IV	2.14		.168	.328 ~	.267	2.82	.051	3, 39	.178
V	2.65	–.163 *	.232 *			**4.31**	**.020**	2, 40	.177
VI	2.57	–.139	.185 ~	.241	.258	**2.86**	**.036**	4, 38	.231

Note: ~ p≤.10; * p≤.05; ** p≤.01; *** p≤.001; **** p≤.0001

The next to last of these models—Model V—is best not only because it is statistically significant overall, but also because all the coefficients are also statistically significant.[4] This means that together, Initial SWL and Initial LOC predict Initial SOI better than each does separately and better than they do while also controlling for site specific differences in SOI. The equation for this model is

$$InitialSOI = 2.65 + (-.163) \times InitialSWL + .232 \times InitialLOC$$

This model says that higher LOC scores are associated with higher SOI scores at constant levels of SWL, but also that higher SWL scores are associated with *lower* SOI scores at constant levels of LOC. Although it is consistent with constructive-developmental theory to say higher SOI scores are associated with higher Locus of Control scores—the SOI measures, in part, where a person locates authority—what does it mean to say that higher SOI is associated with lower scores on the

[4] A comparison of Models V and VI shows that jointly adding the site related variables yields a model (VI) that is not statistically significantly different from V, which does not include these variables (F=1.33, df=2,38, p=.28).

Satisfaction with Life scale? It could be that this finding is also an artifact of the particular distributions within our populations, especially as we found this relationship is strongest within the Polaroid site.[5]

Relationships with Final SOI

We can further explore the robustness of these relationships by trying to find similar patterns at our final data collection visit. Although it did not make sense to predict Initial SOI from scores at our final data collection visit, the converse is not true for Final SOI scores—looking for predictors of Final SOI from the initial visit is consistent with our theoretical approach. Not surprisingly, we find Initial SOI to be a very strong predictor of Final SOI (Model VII in Table 6). Constructive-developmental level does not, prior research suggests, change much over the relatively short timeframes studied here.

As we investigated other such relationships, we found one person had anomalous scores at our final data collection visit, masking relationships otherwise apparent within the data. This participant had the lowest Locus of Control (LOC) (2.3) and change in LOC (-2.2) scores, and the second lowest Efficacy (PEBS) score (3.3) and the lowest change in PEBS score (-1.2). Relationships between Final SOI and LOC are not statistically significant if we include her but they are significant if we exclude her. Thus, in our analyses, we have chosen to exclude her anomalous data.[6]

Table 6: Predicting Final SOI by Other Psychological Variables

Model	Intercept	SOI Initial	LOC Final	F statistic	p	df	R^2
VII	.30	.912 ****		**167.22**	**.0001**	1, 36	.823
VIII	1.93		.256 *	**6.14**	**.020**	1, 27	.185

Note: ~ p≤.10; * p≤.05; ** p≤.01; *** p≤.001; **** p≤.0001

Other models predicting Final SOI by SWL, PEBS, and Site individually or in combination with each other and LOC are not statistically significantly different from the null hypothesis that all parameter estimates are 0. (i.e., they are no better than simply using average values for predicting SOI scores.) Notice that the parameter estimates for Model VIII are similar to those for Models III and V and state, again, that higher LOC scores are associated with higher SOI scores at our final data collection visit. Interestingly, the difficult-to-explain negative association with Satisfaction with Life that we found at our initial visit no longer holds true.

We also explored relationships between changes over time in students' scores on the paper and pencil measures and changes in SOI scores and found no statistically significant relationships. Thus, even though a higher LOC score is associated with higher SOI scores at both data collection visits, *increases* in LOC over time are not associated with *increases* in SOI. This may mostly be

[5] How do we make sense of this finding? The relationships identified here are not causal. It is not that higher levels of cognitive complexity do not lead to lower life satisfaction, or vice versa. At the same time, one could invent a rationale for either of these claims—e.g., that lower life satisfaction drives people towards developmental change as measured by the SOI; or that those with higher cognitive complexity - especially in an environment that does not particularly support it - can find themselves less satisfied with their lives. Such hypotheses, while provocative, would need further investigation and exploration to be sustained.

[6] In the main text, we explore sources of the unusual difference in this learner's experiences.

because SOI changed so little, though it may also be that LOC does not change smoothly in the process of development.

In the end, we found a strong relationship between measures of SOI across time, as expected by Kegan's theory, and a consistent relationship between SOI and LOC. Because the relationship between SOI and SWL does not stand up over time, we can cautiously rule it out as an aberrant finding.

Psychological Variables Predicting Each Other

We then examined relationships among the psychological variables measured by the paper and pencil measures. One interesting general finding was that scores on any particular measure from our initial data collection visit do not predict scores on the same measure for our final data collection visit. People seem to change– and change in both directions– on these measures.

However, *Change* in each of the paper and pencil measures is strongly negatively predicted by initial values, as shown in Table 7. For LOC and PEBS, analyses are shown both with and without the aberrant data of one participant. Understanding the relationship described here is complex. Although a regression equation with a negative coefficient for the main effect (as seen in all of these examples) might seem to imply that those with higher scores initially were likely to have their scores go down and vice versa, factoring in the intercept gives a slightly different picture.

Table 7: Predicting Change in Psychological Variables by Initial Values

Model	Dependent variable	Intercept	SWL Initial	LOC Initial	PEBS Initial	F statistic	p	df	R^2	Notes
IX	Δ SWL	3.41	–.928 ***			**35.80**	**.0001**	1, 27	.570	
X	Δ LOC	3.39		–.917 ***		**27.45**	**.0001**	1, 25	.523	
XI	Δ LOC	3.00		–.792 ***		**24.38**	**.0001**	1, 24	.504	No Ak
XII	Δ PEBS	3.00			–.729 ***	**15.04**	**.0007**	1, 25	.376	
XIII	Δ PEBS	2.76			–.657 **	**12.95**	**.0014**	1, 24	.351	No Ak

Note: ~ p≤.10; * p≤.05; ** p≤.01; *** p≤.001; **** p≤.0001

For example, take the equation concerning Change in SWL scores:

$$\Delta SWL = 3.41 - .928 \times InitialSWL$$

We can use this formula to calculate predicted values for Δ SWL at different values of Initial SWL. In this data set, the minimum Initial SWL is 1.5, and at this value the average change in SWL is predicted to be an increase of 2.0 points. For each additional point of Initial SWL score, the average *increase* in SWL will be .93 points smaller. At an Initial SWL of 3.7, the predicted increase in SWL would be zero and at higher values of Initial SWL predicted change in SWL would be negative. When Initial SWL is at its maximum of 5, SWL is predicted to decrease by 1.2 points. So,

the average values predicted from the initial range of 1 to 5 is the much smaller range of 3.5 to 3.8, and at its average initial value (3.26), SWL is predicted to increase by .38.

Thus what we see here is a combination of an overall average increase in SWL, combined with a robust "regression to the mean" whereby, when things vary randomly, unusually high or low values are likely to be more middling the next time they're measured. A similar pattern can be seen with each of the other change variables.

Multiple regression analyses were also conducted to examine whether these relationships varied by site. No statistically significant relationship was found between Site and Δ SWL or Δ LOC, nor did adding a set of Site variables contribute to the predictive power of Models IX through XI. On the other hand, Δ PEBS can be predicted by Site at a nearly significant level (Model XIV in Table 8), and thus we also examined whether adding variables for Site to the Initial PEBS variable could make a more robust model (compare Models XV and XVI with Models XII and XIII).

Table 8: Predicting Change in Efficacy by Site

Model	Dependent variable	Intercept	PEBS Initial	BHCC	EVST	F statistic	p	df	R^2	Notes
XIV	Δ PEBS	.500		–.650 *	–.375	3.00	.069	2, 23	.207	No Ak
XV	Δ PEBS	2.92	–.640 **	–.465 ~	–.377	**6.86**	**.002**	3, 23	.472	
XVI	Δ PEBS	2.66	–.573 **	–.484 *	–.281	**6.24**	**.003**	3, 22	.460	No Ak

Note: ~ p≤.10; * p≤.05; ** p≤.01; *** p≤.001; **** p≤.0001

Though it appears the coefficient for the BHCC variable is statistically significant, the Site variable necessarily involves the addition of both the BHCC and EVST variables, and conducting a hypothesis test that checks the significance of adding both of them together to the model finds that their contribution is not statistically significantly different from the null hypothesis that these coefficients are zero (0).[7] Thus, our best models for predicting Δ PEBS are Models XII and XIII.

Finally, we examined relationships among different paper and pencil measures and found a few interesting results—one among variables at our initial data collection visit, and several at our final data collection visit.

Specifically, we found that Initial LOC scores are predicted by Initial PEBS scores. Higher PEBS scores are associated with higher LOC scores, with coefficients described in Table 9. However, no other variables show a statistically significant relationship at our initial data collection visit, nor does the addition of Site variables change these results.

[7] The ΔF test comparing Models XII and XV yields F=2.10, df=2,23, p=.145. The ΔF test comparing models XIII and XVI yields F=2.23, df=2, 22, p=.132.

Table 9: Regression Coefficients for Relationships Among Different P&P Measures at Initial Visit

Model	Dependent variable	Intercept	PEBS Initial	F statistic	p	R^2	df	Notes
XVII	Initial LOC	1.27	.58 ***	**14.77**	**.0004**	.247	1, 45	

At our final visit, we find relationships among all three of the paper and pencil measures (see Table 10). This is true whether or not our anomalous participant's scores are included, though her scores– especially on the PEBS– are highly influential. PEBS is less highly correlated with both LOC and SWL when her scores are excluded.

Table 10: Regression Coefficients for Relationships Among Different P&P Measures at Final Visit

Model	Dependent variable	Intercept	SWL Final	PEBS Final	F statistic	p	R^2	df	Notes
XVIII	Final LOC	2.06	.449 **		**7.64**	**.010**	.209	1, 29	All
XIX	Final LOC	2.28	.400 *		**6.73**	**.015**	.194	1, 28	No Ak
XX	Final LOC	1.60		.516 *	**6.53**	**.016**	.189	1, 28	All
XXI	Final LOC	2.05		.415 *	**4.25**	**.049**	.136	1, 27	No Ak
XXII	Final LOC	.839	.352 *	.381 ~	**5.93**	**.007**	.305	2, 27	All
XXIII	Final LOC	.874	.334 *	.295	**4.57**	**.020**	.260	2, 26	No Ak
XXIV	Final SWL	2.14		.383 ~	3.12	.088	.100	1, 28	All
XXV	Final SWL	2.25		.362	2.49	.126	.085	1, 27	No Ak

Final LOC scores are predicted by both final SWL scores and final PEBS scores alone. Higher scores on either of these other measures are associated with higher LOC scores, and this is true whether or not our anomalous participant's scores are included. However, Models XXII and XXIII, which try to incorporate both variables into a multiple regression model, show mixed results. Conducting a hypothesis test to determine the power of adding the PEBS variable to models that only contain the SWL variable (comparing model XXII to model XVIII with all the data, and comparing model XXIII to model XIX excluding the anomalous participant), we find that the addition of the PEBS variable is borderline significant with all the data ($F=3.71$, $df=1, 27$, $p=.065$) but not statistically significant if she is excluded ($F=2.30$, $df=1, 26$, $p=.14$). On the other hand, adding the SWL variable to the models that only include PEBS (comparing model XXII to XX, and model XXIII to XXI) yields a statistically significant increase whether including her ($F=4.51$, $df=1,27$, $p=.043$) or not ($F=4.35$, $df=1,26$, $p=.047$). These relationships could be explained if SWL and PEBS are correlated with one another because the variance that each explains would be at least partly explained by the other. The nearly statistically significant relationships found in Models XXIV and XXV confirm this.

Finally, we find a relationship between Δ LOC and Δ PEBS, shown in Table 11 below.

Table 11: Relationship Between Changes in P&P Measures

Model	Dependent variable	Intercept	Δ PEBS	F statistic	p	R^2	df	Notes
XVIII	Δ LOC	.094	.663 *	**7.48**	**.012**	.245	1, 23	All

Taking the results from Tables 9 through 11 together, we find a consistent relationship between LOC scores and PEBS scores—they are correlated at both our initial and final visits, as are changes in both these variables. The relationship between LOC and SWL is strong at our final visit, but not after controlling for PEBS scores, nor does it hold true across visits. Of course, because these relationships are correlational rather than causal, we could have used LOC to predict PEBS or SWL instead of the other way around. We have chosen to report these relationships this way both because the set of correlations are stronger with LOC and because LOC is, in turn, correlated with SOI, which is not true about the other psychological variables.

Thus, across data collection visits, Efficacy Beliefs seem to be related to Locus of Control, and these in turn are related to constructive-developmental level of mind, as measured by the SOI, though Efficacy Beliefs do not predict SOI directly. The relationship between Satisfaction with Life and the other psychological variables is mixed, arising at some data collection visits and not others, showing up in directions that are hard to explain theoretically, and often not strong enough to be considered statistically significant.

Summary and Conclusions

After conducting a reliability analysis for the three psychologically oriented paper and pencil measures we used in this study, we examined both site-specific differences in demographic and psychological variables and relationships among these. We found differences among our sites in demographic characteristics such as Age, Years in the United States, Years of Own and Mother's Education, and Parental Status and Number of Children. We also found a small number of differences among our sites in the psychological variables, specifically seeing slightly higher increases in both Efficacy Beliefs and SOI at Polaroid than at the other two sites.

In looking for relationships among these variables, we found that, of the demographic variables, only Mother's Education predicted SOI. Of the psychological variables, Locus of Control consistently predicts SOI (as does previous SOI scores), and we find inconsistent and confusing relationships between Satisfaction with Life and SOI.

Examining relationships among the paper and pencil measures, we find that Efficacy Beliefs consistently predict Locus of Control scores and more confusing and inconsistent relationships between Locus of Control and Satisfaction with Life scores.

REFERENCES

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71-75.

Lahey, L., Souvaine, E., Kegan, R., Goodman, R., & Felix, S. (1988). *A guide to the subject–object interview: Its administration and interpretation.* Unpublished manuscript.

Pavot, W., Diener, E., Colvin, C. R., & Sandvik, E. (1991). Further validation of the satisfaction with life scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment, 57*(1), 149–161.

Riggs, M. L., Warka, J., Babasa, B., Betancourt, R., & Hooker, S. (1994). Development and validation of self-efficacy and outcome expectancy scales for job-related applications. *Educational and Psychological Measurement, 54*(3), 793–802.