

The Year of THINK 2005-06:

Assessing the THINK Competency for Valencia Community College

Note: Throughout this document, "THINK" refers to Valencia's Student Core Competency: Think clearly, critically, and creatively, analyze, synthesize, integrate and evaluate in many domains of human inquiry.

The objective of this project is to investigate an assessment process for Valencia's core competency of THINK that is ***grounded in the evaluation of actual student work***. This report includes (1) a discussion of the methodology employed to establish the reliability and validity of our process and instruments, (2) a detailed listing of analysis and results, and (3) recommendations for the next Year of THINK and the future assessment of Valencia's core competencies.

Executive Summary

As this was our first foray into the institutional assessment of our core competencies, the investigation focused more on process than product. The reliability and validity analyses of the analytic and holistic THINK rubrics have provided valuable insights into the strengths and weaknesses in our instruments and processes. The extensive process that was used to initially develop the THINK indicators and then integrate them into our instruments resulted in a set of criteria to evaluate THINK that can be widely accepted as reliable and valid for our institution. However, the process involving interdisciplinary teams for holistic THINK scoring can be significantly improved through professional development to increase the inter-rater reliability among the coders. If the inter-rater reliability cannot be improved to acceptable levels in the next round, the institution may want to investigate discipline-specific scoring teams as an alternate plan.

There is strong evidence for the face validity of the rubrics through the many inspections done by Valencia faculty, staff, and administration. Correlations between scores of individual faculty evaluating their own students and the rubric scores of the interdisciplinary teams further strengthen the argument for validity. Weaknesses in our position arise from the lack of external measures of thinking ability to correlate with our own internal scores. This is a weakness we will investigate in future iterations of the assessment of THINK. Finally, while there is no evidence to indicate the THINK instruments measure student abilities other than those specified in the THINK competency, we are recommending this be investigated further in the next year of THINK. Although this is an important issue, we do not believe that the fact that it is unresolved at this time invalidates the use of our THINK instruments.

Our investigation collected student work from 23 individual classes and more than 500 students. One hundred thirty-seven individual pieces of student work were chosen to be scored both analytically and holistically. Class averages ranged from 1.38 to 2.70. These scores indicate the classes ranged from "Beginning" to "Competent" on our THINK levels. The progression of scores correlated to the increase in course level from college preparatory to college level.

As with any process, the initial implementation reveals areas for improvement. The following are recommendations for future assessment of the core competencies.

Recommendations

- It is essential that future instruments for the assessment of the core competencies receive the same thorough attention in their development as the THINK rubrics. This process has already begun for the COMMUNICATE instruments.
- A more in-depth professional development experience with the evaluation criteria and range-finding practice should significantly improve our interdisciplinary holistic scoring processes.
- Some of the student work we received could not be easily scored holistically by our interdisciplinary team. The reasons that the work was excluded should be clearly explained to faculty participants in future assessment projects so they understand what we can, and cannot, score holistically.
- It is vitally important that all relevant data be collected during the implementation of the project. This should include data to establish reliability and validity as well as actual assessment results. (See recommendation #4 in Part 3: Recommendations for specific data to collect.)
- Weaknesses in our validity analysis arise from the lack of external measures of thinking ability to correlate with our own internal scores. This is a weakness we hope to remedy in future iterations of the assessment of THINK and the other core competencies.

PART 1: METHODOLOGY

Through an extensive process of reviewing literature and consulting with leaders in the field of the institutional assessment of collegewide learning outcomes, the Learning Evidence Team (LET) developed a plan to use rubrics to assess Valencia's THINK competency. (See "Learning Evidence Plan for Transformative Assessment" <http://www.valenciacc.edu/learningevidence/products.asp> for more background information.) Two rubrics, both based on the collegewide THINK indicators, were developed with different purposes in mind. First, an analytic rubric was developed for individual faculty members to use in their classes to evaluate their students' mastery of THINK and provide specific feedback to students on their strengths and weaknesses in this competency. Second, a holistic rubric was developed for the institution to use as a tool to assess THINK outside of individual classrooms and across departments and programs in order to get an annual snapshot of the institution as a whole. (See Appendix A for copies of both rubrics.) This paper discusses the methodology used to establish the reliability and validity of these two instruments as measures of THINK for each of their intended purposes. A short description of the development of both rubrics is included in the validity analysis but more information is available at <http://www.valenciacc.edu/learningevidence/>.

Reliability

Reliability is defined as the extent to which the assessment results can be replicated and yield consistent results. There are several measures used to establish reliability. The type of reliability measure needed for a particular instrument is determined by the nature and use of that instrument. As the analytic and holistic rubrics are inherently different in how they are intended to be used, different strategies are needed to establish their respective reliabilities. Each of these strategies is discussed individually.

Reliability of the Analytic THINK Rubric

The analytic THINK rubric presents the THINK indicators and levels of achievement in a matrix form which provides for the assessment of each indicator individually. It is the intent of the LET that faculty from across the disciplines use the analytic THINK rubric to assess their students individually and that the results be shared and discussed with their students to provide both formative and summative feedback on the students' abilities to THINK. The five indicators, or criteria, listed on the analytic rubric are all intended to measure students' abilities to THINK as defined by Valencia's Student Core Competencies. (See "Student Core Competencies" <http://valenciacc.edu/competencies/default.asp>.) Thus, an analysis of the reliability of this instrument to assess THINK must investigate whether or not these five indicators were consistently used by faculty from across the disciplines in a manner which demonstrates that all the indicators are measuring the same concept. Simply stated, we need to find out if all five of the indicators belong on the same rubric.

Cronbach's alpha is a statistical procedure that calculates how well a set of items measures a single construct, such as critical thinking. It helps us determine if the five different indicators on our analytic THINK rubric are related closely enough to each other to conclude that they are all being used to measure the same thing (i.e., THINK). The value of alpha is always between zero and one, and a value of 0.70 or higher is considered "acceptable" to establish internal reliability in most Social Science research situations, although values of 0.80 and above are preferred.

THINK projects were completed for 23 individual classes. Instructors of these classes submitted class sets of student work, which were designed for students to demonstrate THINK and then evaluated by the classroom instructor using the analytic THINK rubric. Of the original 23, we had to eliminate eight in the analysis because of incomplete data, small sample sizes, or other anomalies. Our analysis for the remaining 15 classes, which represented 10 different disciplines and 271 students, revealed that all sections had significant and acceptable levels of internal reliability. (See Part 2: THINK Results Report for complete listing of calculations of Cronbach's alpha by course.) Although there are courses at the institution that are not represented by this project, the diversity of courses that are represented and their consistently strong measures of alpha provides evidence that the analytic THINK rubric is a reliable instrument for Valencia faculty members to use for the evaluation THINK regardless of discipline or program.

Reliability of the Holistic THINK Rubric

Unlike the analytic rubric, the holistic THINK rubric presents the THINK indicators and levels of achievement in a form that does not allow each indicator to be assessed separately. Instead, a single judgment is made for the entire piece of student work based on how well that work demonstrates the indicators collectively. It is the intent of the LET that the holistic THINK rubric be used as a summative measure for department, program, and even institutional assessment of students' mastery of THINK. The scoring in this assessment project was done by an interdisciplinary team of faculty. Each piece of student work was scored by two different faculty members. Thus, an analysis of the reliability of this instrument requires us to evaluate the consistency in scoring by the different readers, also known as inter-rater reliability. It is important to note that although the holistic rubric functions differently than the analytic rubric, it uses the same indicators to evaluate THINK as the analytic rubric. The positive reliability results from the analytic rubric analysis help support our argument for the internal reliability of the holistic rubric as well.

Cohen's kappa is a calculation used to determine inter-rater reliability, or the consistency between two different coders of the same piece of student work. Calculating kappa helps uncover whether or not holistic scores are dependent on the faculty member scoring the artifact. Kappa assesses the extent to which two readers make the identical decision in assigning a holistic score to the same piece of student work adjusted for the possibility that their scores may match by random chance. For our coders, $\kappa = .66$, which is below an "acceptable" value of 0.70. In reality, our kappa would be even lower, but in this round of training we defined a "match" as matching or contiguous scores. In future iterations of this project, it would be prudent to work toward consistency defined in terms of only exact matches in scores by both readers. These results indicate that,

although our THINK criteria have been shown to be reliable, we need to significantly improve our training for holistic scoring participants to improve inter-rater reliability with interdisciplinary faculty evaluators in the future. The inter-rater reliability measure, kappa, would need to be above 0.70 before this instrument should be used for interdisciplinary assessment at program levels and above. There was not enough data to determine if the holistic THINK rubric has a high enough inter-rater reliability to be used consistently by discipline-specific faculty.

Validity

Establishing validity shows that an instrument or assessment process measures what it is intended to measure. In essence, we need to show that the THINK rubrics really measure THINK. To be valid, the rubrics must be reliable, but reliability alone does not guarantee their validity. To establish the validity of an instrument, four questions must be answered. First, does a review of that instrument by an informed examiner indicate that it appears to be a valid measure? Second, can the results from the instrument be used to predict student performance in specific abilities? Third, are the results of the instrument related to other measures of the same concept? Finally, does the instrument measure only what it is supposed to measure and nothing else? Answering these questions helps us establish face validity, criterion validity, convergent validity, and discriminant validity respectfully.

Face Validity

Does a review of the THINK rubrics indicate that they appear, on the face of it, to be valid measures of THINK? Both rubrics are based upon the THINK indicators, *which were developed by many groups of faculty, staff, and administration over several years to define how Valencia graduates would demonstrate their mastery of the THINK competency*. A work team of several interdisciplinary faculty members drafted the first version of the analytic THINK rubric from the original indicators in the spring of 2005. During this process, what each criterion and level of achievement on the rubric meant to different disciplines was discussed. Beginning in the summer of 2005, the draft analytic rubric was shared with faculty, administration, and the governing councils of the college through a series of meetings and workshops. Some faculty members were asked to create discipline specific THINK assignments for their students to check to see if the rubric could measure THINK for those assignments. This process revealed areas of ambiguity in the rubric criteria. Revisions were made to the rubric to reflect concerns and needs for clarification.

Once the analytic rubric reached the point where there were no more significant requests for revision, the criteria from the analytic rubric were crafted into the overall levels of achievement on the holistic rubric. The holistic rubric went through a similar process of sharing and feedback. In the spring of 2006, more than 30 interdisciplinary faculty members used the analytic and holistic rubrics to evaluate their students' mastery of THINK for an assignment that they developed for their classes. No significant revisions to the rubrics were recommended by any of these faculty members. An exact count of all those who inspected and used both of the rubrics in the 2005-06 THINK

project is impossible, but the number is easily in excess of 100. We believe that using the existing THINK indicators that Valencia educators articulated for student competence in THINK and the amount of input provided by such a large number of Valencia educators contributed to the fact that the last group to use the rubrics had no significant suggestions for change. This provides enough evidence to establish the face validity of the analytic and holistic rubrics.

Criterion Validity

Can the results from the THINK rubrics be used to predict student performance in specified abilities? The answer to this question depends on which version of the rubric we analyze. Recognizing that the analytic rubric is intended primarily for the use of individual faculty to evaluate their own students' abilities to THINK, we asked faculty to discuss in surveys and focus groups how well the rubric worked for this purpose. Feedback from the faculty members who used the analytic rubric to evaluate their own students' performances on specific class assignments points to a positive conclusion. Faculty responses to survey and focus group questions on their use of the analytic rubric reflect a positive correlation between faculty assessment of the students' mastery of THINK via the rubric and the grades students received on the THINK assignments. One faculty member put it this way:

“I decided to give a grade for content and an acknowledgment to the level of critical thinking. I found that, for the most part, the level of critical thinking matched with the level of content.” – Math Instructor

Thus, higher scores on THINK rubric criteria could be called a predictor of higher grades on the class assignments.

Pearson's correlation is the value normally calculated to determine the strength of the relationship between two variables, or how well one variable can predict another. However, our project did not include the collection of the actual class grades students received on the THINK assignments so we cannot calculate the specific correlations between the rubric scores and the classroom grades. We can only substantiate criterion validity for the analytic rubric through the survey and focus group feedback we have from faculty. In future iterations of this project we will collect the actual classroom grades to allow for the calculation of the specific correlations to strengthen our position.

The holistic THINK rubric was used by both individual faculty members and an interdisciplinary team of faculty to evaluate the same student work. The holistic scores from the individual faculty members and interdisciplinary team members were correlated to determine the strength of the predictive relationship between them. All seven courses that were evaluated showed a statistically significant correlation. However, in two of the seven courses the holistic team score would only be considered a “moderately weak” predictor of the classroom instructor's score. (See Part 2: THINK Results Report for the complete listing of correlations by course.) These mixed results leave some question as to the ability of the interdisciplinary team's holistic scores to predict student performance as evaluated by the classroom instructor. However, the holistic team score was also able to discriminate the level of the course that produced the student work. This means that college preparatory and lower level coursework received lower holistic team scores and upper level coursework received the higher holistic team scores. Together, these positive

correlations form our argument for the criterion validity of the holistic THINK rubric. This argument will be strengthened in the future by the collection of more information from a wider variety of classes and disciplines.

Convergent Validity

To establish convergent validity, we need to show that the results of the holistic THINK rubric can be correlated to another measure of the same construct. This is problematic because the THINK rubrics were designed to measure THINK as it is defined for an internal Valencia population. Valencia's definition of THINK is not necessarily equivalent to other existing concepts of critical thinking and their corresponding measurement instruments. Thus, we cannot at this time correlate our holistic scores with another measurement of THINK external to our institution.

The lack of an external THINK measurement to correlate with our holistic findings turned our focus internal. All of the participants in our THINK projects have received professional development in Valencia's Student Core Competencies and are very familiar with how the THINK competency is defined at our institution. Therefore, we decided to use the same correlation calculations, detailed above in the description of criterion validity that measure the strength of the relationship between the holistic THINK scores from the interdisciplinary team and those that were assigned by the classroom instructors, to validate the holistic rubric. The scores from some classes had stronger relationships than others, but the results for all the class comparisons indicate a statistically significant, positive relationship between the interdisciplinary team's and classroom instructor's holistic scores for THINK. (See Part 2: THINK Results Report for detailed listing of correlations.) Future iterations of our assessment of THINK at Valencia will seek to find an external critical thinking instrument that we can use to better establish the convergent validity of the holistic rubric.

Discriminant Validity

Our final validity question investigates whether or not the THINK rubrics measure only what they are supposed to measure and nothing else. This question reveals the greatest area of weakness in our validity analysis. Whereas criterion and convergent validity are evidenced by strong correlations between measures, discriminant validity is usually evidenced by weak correlations with measures that evaluate something other than THINK. It can be argued that THINK is a construct that is part of many other skill sets that students could be asked to demonstrate, and thus, THINK scores would have some relationship with the scores from any of these evaluations. At this time, we have not found a procedure to adequately establish the discriminant validity of the THINK rubrics. However, it is noteworthy to indicate that members of the interdisciplinary scoring group did make comments such as "If the grammar or sentence structure were taken into account rather than just the THINK indicators, the student artifact would have received a lower holistic score." This is an indication that there was an effort to distinguish the evaluation of THINK from other skills. It is also important to note that ongoing discussions in professional development programs at the institution reveal that many

faculty believe that THINK cannot be measured discriminantly from other abilities and that this type of validity cannot be established for any instrument.

Conclusion

This reliability and validity analysis of the analytic and holistic THINK rubrics has provided valuable insights into the strengths and weaknesses in our instruments and processes. The extensive process that was used to initially develop the THINK indicators and then integrate them into our instruments resulted in a set of criteria to evaluate THINK that can be widely accepted as reliable and valid for our institution. However, the process involving interdisciplinary teams for holistic THINK scoring can be significantly improved through professional development to increase the inter-rater reliability among the coders. If the inter-rater reliability cannot be improved to acceptable levels in the next round, the institution may want to investigate discipline-specific scoring teams as an alternate plan.

There is strong evidence for the face validity of the rubrics through the many inspections done by Valencia faculty, staff, and administration. Correlations between scores of individual faculty evaluating their own students and the rubric scores of the interdisciplinary teams further strengthen the argument for validity. Weaknesses in our position arise from the lack of external measures of thinking ability to correlate with our own internal scores. This is a weakness we hope to remedy in future iterations of the assessment of THINK. Finally, while there is no evidence to indicate the THINK instruments measure student abilities other than those specified in the THINK competency, we are recommending this be investigated further in the next Year of THINK. Although this is an important issue, we do not believe that the fact that it is unresolved at this time invalidates the use of our THINK instruments.

PART 2: THINK RESULTS REPORT

The following report includes various analyses from data collected by the Learning Evidence Team (LET) during the Year of THINK. Short descriptions are included with each set of information. See the Methodology report for detailed descriptions of the calculations and their meanings.

Analytic Rubric Results

There were a total of 23 courses represented in Year of THINK projects. Eight of these courses had to be omitted from analysis because of small sample sizes or missing information. The remaining 15 courses with 271 total students are included in these analyses of results from the analytic rubric.

Mean Scores

The following table includes the average scores for each of the five THINK indicators by course as assigned by the individual classroom instructors. The following are the levels of achievement used to evaluate student work.

1= Beginning 2=Developing 3=Competent 4=Accomplished NA=Not Assessed

Course	Mean Score				
	Analyzing Information	Applying Procedures	Presenting Perspectives	Drawing Conclusions	Synthesizing Ideas
AMH2020	3.40	NA	2.60	3.10	2.90
CHM1025C	2.67	2.52	2.76	2.76	2.81
EAP0440	NA	NA	1.50	1.40	1.70
EAP1520	1.97	1.97	1.57	1.71	1.31
EAP1620	1.76	1.47	1.29	1.24	1.35
ENC1102 (Discussion)	2.05	NA	1.60	1.90	1.90
ENC1102 (Essay)	2.48	2.43	NA	2.38	2.43
HUM2223	3.00	3.00	2.94	2.89	3.00
MAC1105	3.0	2.96	2.78	2.83	3.09
MAT1033	2.25	2.29	2.29	2.0	2.04
OST1335	2.50	2.50	NA	2.50	2.60
REA0001	1.70	2.00	2.10	2.20	2.10
REA0002	2.69	2.69	2.62	2.28	NA
SPC1600	2.44	2.25	1.88	2.44	1.75
SLS	3.43	NA	3.43	3.71	4.00

Internal Reliability

Cronbach's alpha is a measure of reliability or consistency. It determines how well a set of items measures a single construct by measuring the intercorrelation (strength of the relationship) between the criteria on a rubric. It helps determine if the five different criteria on our analytic THINK rubric are related to each other closely enough to conclude that they are all measuring the same thing (i.e., THINK). The value of alpha is always between 0 and 1 and a value of 0.70 or higher is considered "acceptable" for reliability in most Social Science research situations. Note that all of our values for alpha fall in acceptable levels to establish internal reliability.

Course	Sample Size	Number of Criteria Used	Reliability (Alpha)
AMH2020	10	4	0.8429
SPC1600	16	5	0.9048
OST1335	10	4	0.9889
ENC1102 (Discussion)	20	4	0.9066
ENC1102 (Essay)	21	4	0.8693
MAT1033	24	5	0.8861
EAP0440	10	3	0.8819
EAP1620	17	5	0.9338
EAP1520	35	5	0.9056
REA0001	10	5	0.9338
REA0002	29	4	0.8666
CHM1025C	21	5	0.8677
HUM2223	18	5	0.9917
MAC1105	23	5	0.8303
SLS	7	4	0.7805

Holistic Rubric Results

A total of 137 student papers from seven different class sections were scored with the holistic THINK rubric. The papers were scored by the classroom instructor and by an interdisciplinary team of eight faculty members; each paper was read by two members of the team. These scores are included in the following analyses of results from the holistic THINK rubric.

Mean Scores

The following table includes the average holistic scores by course as assigned by the interdisciplinary holistic scoring team. The following are the levels of achievement used to evaluate student work.

1= Beginning 2=Developing 3=Competent 4=Accomplished

It is important to note that the scores in the table are ranked in order from least to greatest and that there is a discernable progression in course level as the scores increase. (Note that the two ENC1102 courses had significantly different assignments. One evaluated an informal discussion forum while the other was a formal essay.) This progression helps to validate our holistic rubric as a tool for measuring student development in the THINK competency.

Course	Sample Size	Holist Mean Score
EAP1620	17	1.38
OST1335	10	1.75
REA0002	29	1.91
SPC1600	16	2.13
ENC1102 (Discussion)	20	2.35
MAT1033	24	2.46
ENC1102 (Essay)	21	2.70

Inter-rater Reliability

Cohen's kappa is the calculation used to determine inter-rater reliability, or the consistency between two different coders of the same piece of student work. Kappa is based partially on the percent of agreement in score by the two readers from the interdisciplinary team and does not include the score given by the classroom instructor. For our readers, kappa= .66, which is below an "acceptable" value of 0.70. This result indicates we need to significantly improve our training for holistic scoring participants to improve inter-rater reliability in the future.

Correlations

Correlation measures the strength of the relationship between two variables. Strong correlations provide evidence for the validity of the holistic THINK rubric. Correlation is always a value between -1 and 1. The closer to zero the correlation is, the weaker the relationship. All of the correlations in the table below are statistically significant. This means the connection between the holistic scores given by the individual faculty members and the interdisciplinary team members are strong enough that it should not have occurred by chance alone. However, two sets of student work, MAT1033 and ENC1102 (Essay), have lower correlations than the others and should be scored again if given the opportunity.

Course	Correlation
EAP1620	0.73
OST1335	0.66
REA0002	0.79
SPC1600	0.83
ENC1102 (Discussion)	0.75
ENC1102 (Essay)	0.48
MAT1033	0.48

PART 3: RECOMMENDATIONS

This investigation uncovered strengths and weaknesses of our efforts to assess Valencia's Student Core Competencies through the analysis of actual student work. The Learning Evidence Team (LET) has the following recommendations for future assessment projects.

General Recommendations for Assessment of the Core Competencies

Recommendation 1: The approach used to develop the analytic and holistic THINK rubrics was extensive and time-consuming. Yet, it was a vital part of the process to build faculty ownership of these instruments and establish their validity. While these processes can be improved upon, it is essential that future instruments for the assessment of the core competencies receive the same attention in their development as the THINK rubrics and that this process not be abbreviated in any way. (Note: This process is already underway for the COMMUNICATE instruments.)

Recommendation 2: Holistic scoring is a process that many faculty and staff at our institution are unfamiliar with. Our faculty volunteers participated in a two-hour workshop to familiarize themselves with the rubric criteria followed by a second two-hour workshop on holistic scoring and range finding before they actually scored student work. We discovered that this was not enough experience to adequately prepare the faculty participants to score the student work with the reliability and consistency that we would want for valid assessment projects. The following are suggestions for improving our holistic scoring process.

- A more in-depth professional development experience with the evaluation criteria and range finding practice should significantly improve our holistic scoring processes.
- We defined a “match” for scoring purposes in this holistic round as matching **or** contiguous scores (for example, 2 and 3 would be a match). We should work towards actual matches in score as our process improves.

Recommendation 3: We received student work for analysis from 23 individual classes and more than 500 students, but some of that work could not be scored holistically by our interdisciplinary team. Bulleted below are some of the reasons why student work had to be excluded. These reasons should be conveyed to faculty participants in future assessment projects so they understand what we can, and cannot, score holistically.

- Work that was completely quantitative or formulaic in nature, such as from mathematics, chemistry, accounting, etc, that did not also have a verbal description of what was demonstrated could not be understood clearly enough by the interdisciplinary team to be scored accurately.
- Work that had written comments on it from faculty members or other students could not be included because those comments might bias the holistic score given by the interdisciplinary team members.

- The team could only score a final summative product, not the drafts and worksheets that led up to that final product.
- Work that was scored as only right or wrong, such as multiple choice or fill in the blank exercises, could not be scored holistically.
- Only work from individual students, not group projects, was scored holistically.
- Work that did not address at least four of the five THINK indicators was not scored holistically. Thus, future holistic scoring projects should encourage faculty to develop student assignments that address all of the indicators for that competency.

Recommendation 4: We realized while in the middle of our assessment process that we did not collect all of the information we needed from our faculty participants and their students. Requesting more information from our participants after their participation in the project had ended was difficult or at times impossible. We recommend that future iterations take time to envision the data to be collected so it is planned well in advance of the actual collection of student artifacts. It is vitally important that all relevant data be collected during the implementation of the project. This should include data to establish reliability and validity as well as actual assessment results. The following are some suggestions for data to collect.

- Each individual piece of student work should have the following: analytic criteria scores from the classroom instructor, holistic score from the classroom instructor, class grade (in percent form) assigned to the student work by the classroom instructor, holistic score from each of the coders from the interdisciplinary team.
- Our validity argument will be strengthened if we can correlate our analytic or holistic scores to an external measure generated from outside the college to establish whether or not our definition of THINK can be transferable between different colleges. We should also look to correlate our scores with other existing measures generated within the college. If possible, the same students who contribute the work should be evaluated by another instrument and those scores should be correlated with the holistic or analytic scores. For example, the COMMUNICATE scores could be correlated with exit test scores from communications classes or other standardized tests.
- Results are strengthened when the pool of student work comes from a diverse set of classes. Efforts should be made to collect work from a variety of courses and programs to ensure all levels of the rubric indicators have been evaluated and are appropriate for any faculty member to use regardless of the discipline or program.

Recommendations for the Next Iteration of THINK

Recommendation 5: The process involving interdisciplinary teams for holistic THINK scoring can be significantly improved through professional development to increase the inter-rater reliability among the coders. If the inter-rater reliability cannot be improved to acceptable levels ($\kappa = 0.70$ or higher) in the next round, the institution may want to investigate discipline-specific scoring teams as an alternate plan.

Recommendation 6: Weaknesses in our validity analysis arise from the lack of external measures of thinking ability to correlate with our own internal scores. This is a weakness we hope to remedy in future iterations of the assessment of THINK . Also, we have yet to determine if our THINK instruments measure student abilities other than those specified in the THINK competency. While there is no evidence of this, we are recommending this issue be investigated further in the next year of THINK.

APPENDIX A
HOLISTIC AND ANALYTIC THINK RUBRICS

Rubric for the Holistic Assessment of Critical Thinking across the Curriculum

© Valencia Community College Version June 10, 2005

4 Accomplished (precise, insightful, balanced, perceptive, and unified)

Does all or almost all of the following:

- Interprets information (data, ideas, or concepts) accurately, appropriately, and in-depth in new contexts
- Employs formulas, procedures, principles, or themes accurately, appropriately, and/or creatively in new contexts
- Explain—accurately and thoroughly—multiple solutions, positions, or perspectives that balance opposing points of view
- Creates a detailed conclusion or complex solution that is complete, well-supported, logically consistent, and often unique
- Integrates ideas or develops solutions that are exceptionally clear, coherent, and cohesive

3 Competent (accurate, relevant, multiplistic, logical, coherent)

Does many or most of the following:

- Presents information (data, ideas, or concepts) accurately and appropriately in familiar contexts
- Applies formulas, procedures, principles, or themes accurately and appropriately in familiar contexts
- Describes two or more solutions, positions, or perspectives accurately
- Organizes a conclusion or solution that is complete, logical, and consistent with evidence presented
- Connects ideas or develops solutions in a clear and coherent order

2 Developing (correct, appropriate, dualistic, reasonable, consistent)

Does many or most of the following:

- Reports information (data, ideas, or concepts) in familiar contexts with minor inaccuracies, irrelevancies, or omissions
- Uses appropriate formulas, procedures, principles, or themes in familiar contexts with only minor inaccuracies
- Identifies simple solutions, over-simplified positions, or perspectives with only minor inaccuracies
- Offers an abbreviated conclusion or simple solution that is mostly consistent with the evidence presented, with minor inconsistencies or omissions
- Arranges ideas or solutions into a simple pattern

1 Beginning (inaccurate, inappropriate, singular, illogical, fragmented)

Does all or almost all of the following:

- Copies information (data, ideas, or concepts) often inaccurately, incompletely, or omits relevant information
- Labels formulas, procedures, principles, or themes inaccurately, inappropriately, or omits them
- Names a single solution, position, or perspective, often inaccurately, or fails to present a solution, position, or perspective
- Attempts a conclusion or solution that is inconsistent with evidence presented, that is illogical, or omits a conclusion or solution altogether
- Lists ideas or expresses solutions in a fragmentary manner, without a clear or coherent order

*This rubric is intended for use in the assessment of student learning and the improvement of instruction at the institutional level. Please send your comments and suggestions about this rubric to Kurt Ewen at kewen@valenciacollege.edu.

© Valencia Community College

Rubric for the Analytical Assessment of Critical Thinking across the Curriculum
 ©Valencia Community College Version June 10, 2005

	Levels of Achievement			
THINK INDICATORS	Beginning	Developing	Competent	Accomplished
Analyzing information: data, ideas, or concepts	<p style="text-align: center;">Inaccurate</p> Copies information (data, ideas, or concepts) often inaccurately, incompletely, or omits relevant information	<p style="text-align: center;">Correct</p> Reports information (data, ideas, or concepts) with minor inaccuracies, irrelevancies, or omissions	<p style="text-align: center;">Accurate</p> Presents information (data, ideas, or concepts) accurately and appropriately in familiar contexts	<p style="text-align: center;">Precise</p> Interprets information (data, ideas, or concepts) accurately, appropriately and in-depth in new contexts
Applying formulas, procedures, principles, or themes	<p style="text-align: center;">Inappropriate</p> Labels formulas, procedures, principles, or themes inappropriately, inaccurately, or omits them	<p style="text-align: center;">Appropriate</p> Uses appropriate formulas, procedures, principles, or themes with minor inaccuracies	<p style="text-align: center;">Relevant</p> Applies formulas, procedures, principles, or themes appropriately and accurately in familiar contexts	<p style="text-align: center;">Insightful</p> Employs formulas, procedures, principles, or themes accurately, appropriately and/or creatively in new contexts
Presenting multiple solutions, positions or perspectives	<p style="text-align: center;">Singular</p> Names a single solution, position, or perspective, often inaccurately, or fails to present a solution, position or perspective	<p style="text-align: center;">Dualistic</p> Identifies simple solutions, over-simplified positions, or perspectives with minor inaccuracies	<p style="text-align: center;">Multiplistic</p> Describes two or more solutions, positions, or perspectives accurately	<p style="text-align: center;">Balanced</p> Explains—accurately and thoroughly—multiple solutions, positions, or perspectives that balance opposing points of view
Drawing well-supported conclusions	<p style="text-align: center;">Illogical</p> Attempts a conclusion or solution that is inconsistent with evidence presented, that is illogical, or omits a conclusion or solution altogether	<p style="text-align: center;">Reasonable</p> Offers an abbreviated conclusion or simple solution that is mostly consistent with evidence presented, with minor inconsistencies or omissions	<p style="text-align: center;">Logical</p> Organizes a conclusion or solution that is complete, logical, and consistent with evidence presented	<p style="text-align: center;">Perceptive</p> Creates a detailed conclusion or complex solution that is well-supported. logically consistent, complete and often unique
Synthesizing ideas into a coherent whole	<p style="text-align: center;">Fragmented</p> Lists ideas or expresses solutions in a fragmentary manner, without a clear or coherent order	<p style="text-align: center;">Consistent</p> Arranges ideas or solutions into a simple pattern	<p style="text-align: center;">Coherent</p> Connects ideas or develops solutions in a clear and coherent order	<p style="text-align: center;">Unified</p> Integrates ideas or develops solutions that are exceptionally clear, coherent, and cohesive

*This rubric is intended for use in the assessment of student achievement at the institutional level. It can also be used as a guide for development of rubrics to measure critical thinking at the program, course and section levels. Please send your comments and suggestions about this rubric to Kurt Ewen at kewen@valenciacollege.edu