Evaluating a Rubric for Assessing Constraint-Based Solid Models

Dr. Theodore J. Branoff, Illinois State University

Dr. Branoff is a professor and chair of the Department of Technology at Illinois State University. He taught engineering graphics, computer-aided design, descriptive geometry, and instructional design courses in the College of Education at North Carolina State University from 1986-2014. He also worked for Siemens-Switchgear Division and for Measurement Group, Inc. Dr. Branoff's research interests include constraint-based solid modeling strategies and spatial visualization abilities in undergraduate students. He has conducted CAD and Geometric Dimensioning and Tolerancing workshops for both industry and education professionals. Dr. Branoff served as President of the International Society for Geometry and Graphics from 2009-2012. In 2013 he was elected into the Academy of Fellows of the ASEE, and in 2014 he received the Distinguished Service Award from the Engineering Design Graphics Division of ASEE. In April of 2015 Dr. Branoff received the Orthogonal Medal for distinguished service in graphic science from the Technology, Engineering & Design Education faculty at North Carolina State University.

Dr. Kevin L Devine, Illinois State University

Kevin is the Program Coordinator for the Engineering Technology major at Illinois State University. His primary teaching assignments are in engineering graphics, industrial robotics, and CNC programming/machining.

Dr. Josh Brown, Illinois State University

Evaluating a Rubric for Assessing Constraint-Based Solid Models

Abstract

A study was conducted during the Fall 2013 semester to examine the effectiveness of a rubric for evaluating constraint-based solid models. The rubric was created after studying conceptual frameworks and other research related to evaluating constraint-based CAD models. Since only one researcher evaluated the models in the 2013 study, it was recommended that a study be conducted where multiple experts evaluated the same models using this original rubric.

During the Fall 2015 semester, three faculty experts in constraint-based modeling used the same rubric to evaluate a representative sample of models created by students in an introductory engineering graphics course. This paper presents literature related to evaluating constraint-based solid models and inter-rater reliability, describes the methodology and results of the study, and provides recommendations for further research related to evaluating constraint-based solid models.

Introduction / Review of Literature

As the tools for creating virtual models have evolved, engineering graphics educators have continued to adjust their methods for accurately and consistently evaluating students' modeling strategies. Some of these methods include using concise rubrics for evaluating models [1-6], developing activities where students can evaluate their own models [7], and using automated electronic evaluation tools [8-9]. One of the main challenges has been developing a method that clearly informs students about how their models will be evaluated, is a valid and reliable tool for assessing design intent, and allows faculty to evaluate models in a timely and consistent manner.

Rubrics have been shown to provide reliable scoring of performance and have the potential to promote learning and/or improve instruction [10]. The main purpose of the rubric used for evaluating the models in this study was to create a valid and consistent method for scoring constraint-based models used in engineering graphics courses. It was created based on a review of literature of several key topics in engineering graphics, graphicacy and modeling, and constraint-based CAD [4]. These topics include CAD modeling strategies [11-13], conceptual framework research of CAD expertise [14-18], studies related to evaluating CAD models [5-9, 19-20], and engineering graphics literacy [1-4]. Figure 1 displays the main categories of the rubric with an explanation of each.

The rubric described here was used in a previous study to evaluate three models created by 23 students in a second level engineering graphics course [4]. The purpose of that study was to compare this rubric to a more elaborate rubric used to assess engineering graphics literacy [1-3]. Conclusions from this study revealed that scores were significantly higher when evaluated with the new rubric than when evaluated with the older rubric. There were also concerns that the older rubric required a great deal of time to evaluate models. Since only one person evaluated all of the models, it was recommended that a study be conducted to evaluate the inter-rater reliability of multiple raters using the same rubric.

Category	Points
Base/Core Feature correctly identified *For some objects this is clear. For other objects there is some flexibility. The base feature should create a good foundation for modeling the rest of the object in an efficient manner.*	10
Orientation of initial sketch plane *The initial sketch plane is important for establishing the viewing direction of the model and also how the model will be oriented in the assembly. It is also critical for establishing the main symmetry plane for models.*	10
Best model origin *As with the base/core feature, the location of the origin is flexible. It should, however, reflect the design intent of the model. For example, if an object has obvious symmetry, it is a good idea to have the origin in the center of the symmetry. This allows one to take advantage of the defaults planes in the part for establishing symmetry.*	10
Sketches are simple, fully constrained, and reflect appropriate design intent *Are the dimensions given in the problem the ones used in the sketches? If symmetry is needed, is this built into the sketch – omitting dimensions in place of appropriate constraints?*	20
Appropriate feature end conditions *For example, Through-All and Next*	10
Correct application of symmetry/duplication *When key dimensions are modified, do features stay centered per design intent? Is this correctly built into sketches and features?*	10
Accuracy/Complete model *Is the total volume accurate? Are all features taken into consideration?*	20
Modeling strategy efficiency *Does the modeling strategy reflect an economy of features, only using necessary sketches and features? Are features based on the given problem?*	10

Figure 1. The Rubric.

Inter-judge / Inter-rater Reliability

Inter-judge or inter-rater reliability "refers to the degree of agreement between two or more observers/judges with respect to their categorization of n subjects/objects" (p. 669) [21]. When examining the relationships between two judges or raters with interval or ratio data, Spearman's rank order correlation coefficient can be used. When there are more than two judges or raters and the data is interval or ratio, Kendall's coefficient of concordance with the intraclass correlation coefficient can be employed (p. 1053) [22].

Research Questions

The current study was designed to determine whether or not using a rubric for evaluating constraint-based CAD models would produce consistent results between three different raters. The specific research questions were:

1. Is there consistent results for the overall model score between the three raters?
2. Are there consistent results for the individual categories between the three raters?

Methodology

During the Fall 2015 semester, 51 technology students completed the second exam in an introductory engineering graphics course on the 15th day of class. The exam consisted of multiple-choice and matching items used to assess textbook information related to introductory constraint-based modeling. The exam also included two constraint-based modeling activities (Figures 2 & 3). For the two activities, students were asked to model the objects with the given dimensions first. They were given the correct values for the distance between points A & B, the area of the specified face, and the total surface area of the part. They were then asked to modify dimensions 1 & 2 with two new values. Again, they were given the correct values for the distance between points A & B, the area of the specified face, and the total surface area of the part. Finally, they were asked again to modify dimensions 1 & 2 with two new values without being provided the correct distance between points A & B, the correct area of the specified face, or the correct total surface area of the part.

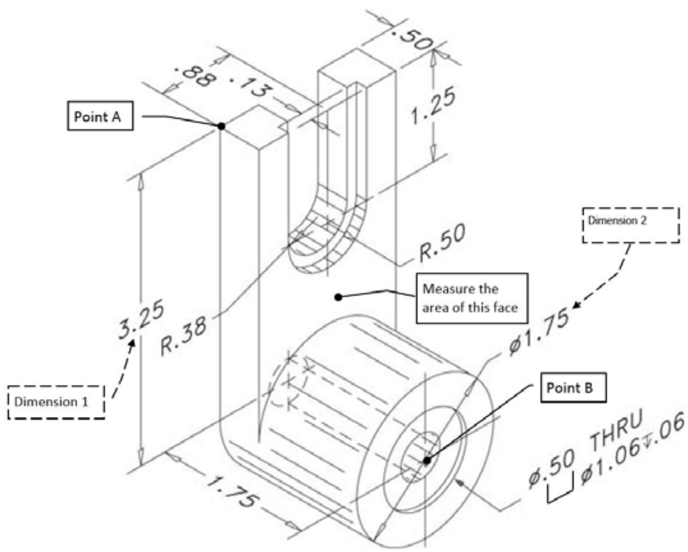

Figure 2. Model 1 [23].

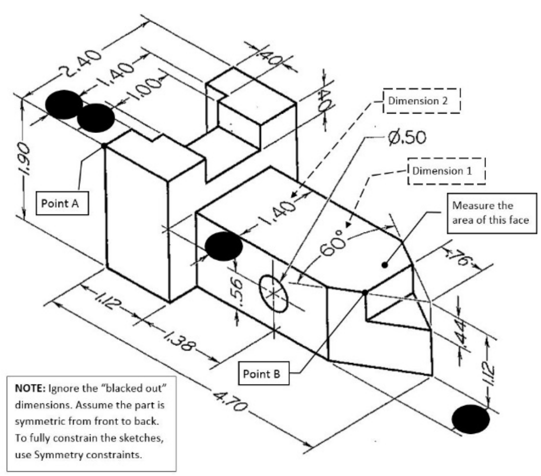

Figure 3. Model 2 [24].

At the end of the semester, one of the researchers evaluated the two exam models completed by the 51 participants using the rubric in Figure 1 (102 total models). A purposeful sample of 10 models was selected for each of the two parts. The samples included 3 models with a score above 90, 3-4 models with scores between 70-90, and 3-4 models with scores below 70. Next, two additional experts in constraint-based modeling independently evaluated the 10 models in the two samples (20 total models). The raters were asked to keep notes on issues related to the rubric as they were evaluating the models. The data were then combined for analysis. Table 1 displays the raw overall scores for the two parts evaluated by the three raters.

Table 1. Raw overall scores for the three raters.

Rater 1 Model 1	Rater 2 Model 1	Rater 3 Model 1	Rater 1 Model 2	Rater 2 Model 2	Rater 3 Model 2
98	94	90	95	77	100
95	95	88	82	82	91
42	62	52	96	83	95
60	81	64	67	62	66
98	96	100	98	94	98
63	65	75	78	83	85
67	57	74	62	52	40
86	75	87	77	75	87
79	75	78	60	70	85
76	68	76	77	83	85

Analysis of Results

Two analyses were conducted to evaluate the reliability of the rubric. A Spearman correlation coefficient was used to look at inter-judge/inter-rater differences (Table 2) [21]. A second analysis examined intra-judge/intra-rater reliability. Since the three raters evaluated the same models and the raters represent similar instructors who may use the rubric to evaluate constraint-based models in their courses, a Two-Way Random Intraclass Correlation (ICC) was used to evaluate the intra-rater effects (Tables 3 & 4). The ICC test takes into account that one rater might rate one model high and another model low, but that the variance should even out across all raters [22].

Table 2. Correlations between raters – Spearman's Correlation Coefficient.

Overall Ratings							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2	* .744	1.000		Rater 2	** .794	1.000	
Rater 3	** .985	** .778	1.000	Rater 3	** .855	.602	1.000
Base/Core feature correctly identified							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2	-.137	1.000		Rater 2	.062	1.000	
Rater 3	* .667	-.365	1.000	Rater 3	-.153	-.445	1.000
Orientation of initial sketch plane							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2	** .976	1.000		Rater 2	** 1.000	1.000	
Rater 3	** .988	** .949	1.000	Rater 3	** .986	** .986	1.000
Best model origin							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2	* .756	1.000		Rater 2	.362	1.000	
Rater 3	* .735	.531	1.000	Rater 3	.550	.239	1.000
Sketches are simple, fully constrained, and reflect appropriate design intent							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2	.483	1.000		Rater 2	.454	1.000	
Rater 3	** .896	.449	1.000	Rater 3	* .730	.295	1.000
Appropriate feature end conditions							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2	** .802	1.000		Rater 2	** .862	1.000	
Rater 3	.675	.449	1.000	Rater 3	-.214	-.166	1.000
Correct application of symmetry/duplication							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2				Rater 2	** .913	1.000	
Rater 3	** .861		1.000	Rater 3	* .671	* .638	1.000
Accuracy/Complete model							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2	** .770	1.000		Rater 2	** .965	1.000	
Rater 3	** .781	.625	1.000	Rater 3	** .990	** .962	1.000
Modeling strategy efficiency							
Model 1				**Model 2**			
	Rater 1	Rater 2	Rater 3		Rater 1	Rater 2	Rater 3
Rater 1	1.000			Rater 1	1.000		
Rater 2	** .912	1.000		Rater 2	* .644	1.000	
Rater 3	** .806	** .878	1.000	Rater 3	* .715	.382	1.000
* Significant at α = .0.05 ** Significant at α = .0.01							

The Spearman's analyses between raters revealed a range of significant correlations. For the overall rating of Model 1 there were significant correlations between all raters. For Model 2, a significant correlation did not exist between raters 2 and 3. The only category within the rubric where significant correlations existed between all raters for both models was orientation of sketch plane. There were significant correlations between all raters for one of the models in the categories of Correct application of symmetry (model 2), Accuracy/complete model (model 2), and Modeling strategy (model 1).

Table 3. Intraclass Correlation Coefficients for Model 1.

Model 1 Ratings		Intraclass Correlations	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	Sig
Overall Rating	Single Measures	.804	.542	.942	13.340	9	18	** .000
	Average Measures	.925	.780	.980	13.340	9	18	** .000
Base/Core Feature	Single Measures	.031	-.264	.504	1.094	9	18	.413
	Average Measures	.086	-1.676	.753	1.094	9	18	.413
Orientation	Single Measures	.942	.841	.984	49.484	9	18	** .000
	Average Measures	.980	.941	.995	49.484	9	18	** .000
Best Model Origin	Single Measures	.604	.232	.868	5.578	9	18	** .001
	Average Measures	.821	.475	.952	5.578	9	18	** .001
Sketch Quality	Single Measures	.604	.231	.867	5.574	9	18	** .001
	Average Measures	.821	.474	.952	5.574	9	18	** .001
Feature End Conditions	Single Measures	.255	-.144	.684	2.027	9	18	.097
	Average Measures	.507	-.445	.867	2.027	9	18	.097
Application of Symmetry	Single Measures	.764	.300	.936	7.477	9	18	** .003
	Average Measures	.866	.462	.967	7.477	9	18	** .003
Accuracy/ Completeness	Single Measures	.677	.331	.896	7.279	9	18	** .000
	Average Measures	.863	.598	.963	7.279	9	18	** .000
Modeling Strategy	Single Measures	.762	.467	.927	10.624	9	18	** .000
	Average Measures	.906	.724	.975	10.624	9	18	** .000
* Significant at α = .05 ** Significant at α = .01								

Table 4. Intraclass Correlation Coefficients for Model 2.

Model 2 Ratings		Intraclass Correlations	95% Confidence Interval		F Test with True Value 0			
			Lower Bound	Upper Bound	Value	df1	df2	Sig
Overall Rating	Single Measures	.766	.473	.929	10.801	9	18	** .000
	Average Measures	.907	.729	.975	10.801	9	18	** .000
Base/Core Feature	Single Measures	-.063	-.315	.405	.822	9	18	.604
	Average Measures	-.216	-2.563	.671	.822	9	18	.604
Orientation	Single Measures	.765	.472	.928	10.772	9	18	** .000
	Average Measures	.907	.728	.975	10.772	9	18	** .000
Best Model Origin	Single Measures	.289	-.088	.706	2.220	9	18	.072
	Average Measures	.550	-.319	.878	2.220	9	18	.072
Sketch Quality	Single Measures	.515	.125	.828	4.182	9	18	** .005
	Average Measures	.761	.300	.935	4.182	9	18	** .005
Feature End Conditions	Single Measures	.049	-.253	.522	1.155	9	18	.377
	Average Measures	.135	-1.535	.766	1.155	9	18	.377
Application of Symmetry	Single Measures	.516	.126	.829	4.198	9	18	** .005
	Average Measures	.762	.302	.936	4.198	9	18	** .005
Accuracy/ Completeness	Single Measures	.955	.875	.988	64.606	9	18	** .000
	Average Measures	.985	.955	.966	64.606	9	18	** .000
Modeling Strategy	Single Measures	.639	.278	.882	6.313	9	18	** .000
	Average Measures	.842	.536	.957	6.313	9	18	** .000
* Significant at α = .0.05 ** Significant at α = .0.01								

The intra-class correlation analyses revealed significant results for the overall evaluations for models 1 & 2 and all but a few of the individual categories within the rubric. For the overall rating of model 1, 92.5% of the variance in the mean of the 3 raters is real variance. For model 2, 90.7% of the variance was real. There was no consensus between raters for the base/core feature and feature end conditions categories for model 1, nor was there consensus between raters for the base/core feature, best model origin, and feature end conditions categories for model 2.

Discussion and Conclusions

This study revealed several interesting results related to using rubrics to evaluate constraint-based CAD models. First, the rubric used to evaluate models in this study appeared to generate reliable results for the overall rating of the two models. The only inter-rater analysis for overall rating that did not result in a significant value was the correlation between raters 2 and 3 for model 2. Looking at the individual categories in the rubric helped to identify specific categories that need refinement. There appears to be quite a bit of variance unaccounted for in the categories of base/core feature, best model origin, and feature end conditions.

The researchers feel inter-rater correlation could possibly be improved through training and/or rating practice models as a group before evaluating the participant models. All three raters were provided the rubric without any further instruction or discussion of the categories. If all raters analyzed practice problems and had group discussions about the rubric, it is possible consensus could be reached on the specific criteria to review for each category of the rubric.

The individual category correlations could be improved by further developing the rubric to include more detailed descriptors of each category. Rubrics can be improved by adding specific criteria expectations for each point of the rubric [25]. The rubric could be modified to include specific criteria for each category on what constitutes a rating at each point value. For example, the category of "Appropriate Feature End Conditions" could include specific criteria for each point value such as a "10" rating includes 100% all end features identified and implemented accurately, a "9" rating includes 90% of all end features identified and implemented accurately, and so on.

Another method to improve the rubric could be to create an "annotated" rubric by listing specific characteristics critical for each category in which the scorer can circle or identify during evaluation [26]. A specific point value is not provided for each characteristic on the annotated rubric, but this would ensure the scorer is looking for specific characteristics of each model evaluated.

The rubric could also be modified to have a consistent scale for each category, such as a 1-5 rating, and then overall weighting of each category could be determined by using a multiplier. Using a consistent scale for each category, instead of variable points, could make evaluating more efficient for the raters.

Future Research

Future research should be conducted on how rubrics can be used in assessing constraint-based solid models. The rubric used in this study could be modified by the previous suggestions and the study repeated. The study could also be repeated with additional raters at other universities. Training or discussion, including sample models, with raters prior to conducting the research could be conducted.

References

1. Branoff, T. J., & Dobelis, M. (January, 2012). Engineering graphics literacy: Measuring students' ability to model objects from assembly drawing information. *Paper published in the proceedings of the 66th Midyear Conference of the Engineering Design Graphics Division of the American Society for Engineering Education, Galveston, Texas, January 22-24, 2012.*
2. Branoff, T. J., & Dobelis, M. (2012). The relationship between spatial visualization ability and students' ability to model 3D objects from engineering assembly drawings. *Engineering Design Graphics Journal*, 76 (3), 37-43.
3. Branoff, T. J., & Dobelis, M. (October, 2013). Spatial visualization ability and students' ability to model objects from engineering assembly drawings. *Paper published in the proceedings of the 68th Midyear Conference of the Engineering Design Graphics Division of the American Society for Engineering Education, Worcester, Massachusetts, October 20-22, 2013.*
4. Branoff, T. J., & Dobelis, M. (June, 2014). Relationship between students' spatial visualization ability and their ability to create 3D constraint-based models from various types of drawings. *Proceedings of the 2014 Annual Conference of the American Society for Engineering Education, Indianapolis, Indiana, June 15-18, 2014.*
5. Ault, H. K., & Fraser, A. (June, 2013). A comparison of manual vs. online grading for solid models. *Proceedings of the 2013 Annual Conference of the American Society for Engineering Education, Atlanta, Georgia, June 23-26, 2013*.
6. Steinhauer, H. M. (2012). Correlation between a student's performance on the Mental Cutting Test and their 3D parametric modeling ability. *Engineering Design Graphics Journal, 76(3),* 44-48.
7. Devine, K. D., & Laingen, M. A. (October, 2013). Assessing design intent in an introductory-level engineering graphics course. *Paper published in the proceedings of the 68th Midyear Conference of the Engineering Design Graphics Division of the American Society for Engineering Education, Worcester, Massachusetts, October 20-22, 2013.*
8. Baxter, D.H. (2003). Evaluating an automatic grading system for an introductory computer aided design course. *Proceedings of the 58th Annual Midyear Conference of the Engineering Design Graphics Division of the American Society for Engineering Education, Scottsdale, Arizona, November 16-19, 2003*.
9. Baxter, D.H. & Guerci, M. J. (2003). Automating an introductory computer aided design course to improve student evaluation. *Proceedings of the 2003 Annual Conference of the American Society for Engineering Education, Nashville, Tennessee, June 22-25, 2003*.
10. Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review 2*. 130-144.
11. Chester, I. (2007). Teaching for CAD expertise. *International Journal of Technology and Design Education, 17,* 23-35.
12. Chester, I. (2008). 3D-CAD: Modern technology – outdated pedagogy? *Design and Technology Education: An International Journal, 12(1),* 7-9.
13. Delahunty, T., Seery, N., & Lynch, R. (2012). An evaluation of the assessment of graphical education at junior cycle in the Irish system. *Design and Technology Education: An International Journal, 17(2),* 9-20.
14. Hartman, N. W. (2003). *Towards the definition and development of expertise in the use of constraint-based CAD tools: Examining practicing professionals* (Unpublished doctoral dissertation). North Carolina State University, Raleigh, North Carolina.
15. Hartman, N. W. (2009). The development of expertise in the use of constraint-based CAD tools: Examining practicing professionals. *Engineering Design Graphics Journal, 68(2),* 14-26.
16. Hartman, N. W. (2009). Defining expertise in the use of constraint-based CAD tools by examining practicing professionals. *Engineering Design Graphics Journal, 69(1),* 6-15.

17. Rynne, A., & Gaughran, W. (June, 2007). Cognitive modelling strategies for optimum design intent in parametric modelling (PM). *Proceedings of the 2007 Annual Meeting of the American Society for Engineering Education,Honolulu, Hawaii, June 24-27, 2007.*

18. Rynne, A., Gaughran, W. F., & Seery, N. (2010). Defining the variables that contribute to developing 3D CAD modelling expertise. In E. Norman & N. Seery (Eds.), *Graphicacy and Modelling*. The International Conference on Design and Technology Educational Research and Curriculum Development, Loughborough, U.K. 161-233.

19. Company, P., Contero, M., & Salvador-Herranz. (July, 2013). Testing rubrics for assessment of quality in CAD modelling. *Proceedings of the Research in Engineering Education Symposium, Kuala Lumpur, Malaysia, July 4-6, 2013*.

20. Peng, X., McGary, P., Johnson, M., Yalvac, B., & Ozturk, E. (2012). Assessing novice CAD model creation and alteration. *Computer-Aided Design & Applications, PACE, (2)*, 9-19.

21. Sheskin, D. J. (2007). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton: Chapman & Hall/CRC.

22. Landers, R. N. (November 16, 2011). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *NeoAcademic*. Accessed January 20, 2016 from: http://neoacademic.com/2011/11/16/computing-intraclass-correlations-icc-as-estimates-of-interrater-reliability-in-spss/

23. Lieu & Sorby, (2009). *Visualization, Modeling, and Graphics for Engineering Design*. Clifton Park, NY: Delmar, Cengage Learning.

24. Giesecke, F. E., Mitchell, A., Spencer, H. C., Hill, I. L., Dygdon, J. T., Novak, J. E., & Lockhart, S. (2009). *Technical Drawing* (13th ed.). Upper Saddle, NJ: Pearson/Prentice-Hall.

25. Miller, M.D., Linn, R.L., & Gronlund, N. (2013). *Measurement and Assessment in Teaching.* Upper Saddle River, NJ: Pearson.

26. Brookhart, S.M. (2008). *How to give effective feedback to your students.* Alexandria, VA:ASCD.