Patterns of Performance Across Different Types of Items Measuring Knowledge of Ohm's Law

Technical Report 405

Brenda Sugrue
University of Northern Colorado

Rosa Valdes, Jonah Schlackman, and Noreen Webb
CRESST/University of California, Los Angeles

March 1996

U.S. Department of Education
Office of Educational Research and Improvement
Grant No. R117G10027 CFDA Catalog No. 84.117G

National Center for Research on
Evaluation, Standards, and Student Testing
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

PATTERNS OF PERFORMANCE ACROSS DIFFERENT TYPES OF ITEMS MEASURING KNOWLEDGE OF OHM'S LAW

Brenda Sugrue, University of Northern Colorado
Rosa Valdes, Jonah, Schlackman, and Noreen Webb, CRESST/UCLA

Abstract

This report presents an analysis of data gathered as part of a larger study designed to investigate the relative contribution of a number of cognitive constructs and test characteristics to variation in performance on a set of items and tasks developed to probe students' knowledge of Ohm's law and how it operates in electric circuits. Specifically, in this report, we consider how interpretations of the nature and extent of students' knowledge might be facilitated by examining patterns of performance within and across different types of items targeting the separate concepts related by Ohm's law: voltage, resistance, and current.

There were consistencies and inconsistencies across and within sets of short written test items. Consistency of performance on short written test items varied by knowledge unit; performance was more consistent across item types for the concept of resistance than for voltage and current. Performance also varied within item type; surface features of the items influenced accuracy. In the context of extended tasks, opportunity to engage in hands-on activity did not affect performance; similar performance levels were observed in written analogues of the hands-on tasks. Within extended tasks, high correlations were found between multiple-choice items and items requiring written explanations for simpler knowledge units, but not for the more complex concept of current.

The results of this study suggest that the test design process should incorporate analysis of patterns of performance within and across sets of items that vary in efficiency and authenticity. Such analysis can guide the selection of a smaller set of items to maximize both the efficiency of the assessment and the validity of the inferences made about students' knowledge and abilities.

PATTERNS OF PERFORMANCE ACROSS DIFFERENT ITEMS MEASURING KNOWLEDGE OF OHM'S LAW

Brenda Sugrue, University of Northern Colorado
Rosa Valdes, Jonah, Schlackman, and Noreen Webb, CRESST/UCLA

Introduction

The recent trend toward alternative assessment has been based on some implicit assumptions about the relative ability of different types of test items to induce and reflect different types and levels of knowledge and cognitive activity. In particular, it is often assumed that ill-defined characteristics of a test, such as complexity, "real-world"-ism, "hands-on"-ness, opportunity to collaborate with others, or opportunity to generate a response as opposed to selecting one from a set of given alternatives, make the test a more valid indicator of the state of a student's knowledge. A number of researchers have begun to question these assumptions (for example, Baker, O'Neil, & Linn, 1993; Baxter, Glaser, & Raghavan, 1993; Snow, 1993), and there is a growing body of research that indicates that the cognitive demands of a task are not always reflected in its surface characteristics (for example, Bridgeman & Rock, 1993; Lukhele, Thissen, & Wainer, 1994; Snow, 1993).

Shavelson, Baxter, and Gao (1993) found that not only did student performance vary across hands-on science tasks with different content; performance also varied across hands-on and analogous tasks in different formats targeting the same content. Bridgeman (1992) found that, although total scores on sets of multiple-choice and open-ended items on the quantitative portion of the Graduate Record Examination were comparable, "at the level of the individual item, there were striking differences between the open-ended . . . and multiple-choice formats" (p. 269). Snow, Ennis, Kupermintz, and Talbert (1994) found that within sets of presumed unidimensional multiple-choice mathematics or science items, patterns of performance across items indicated that different items were actually tapping distinct types of knowledge and reasoning. The apparent sensitivity of judgments about student abilities to variation in the content and format of an assessment raises serious doubts about the validity of all assessments, not just performance-based assessments.

Consequently, researchers are now beginning to use a variety of methods to investigate the cognitive activity induced by both traditional and innovative assessments. The initial goal is to get a better understanding of differences and similarities in the cognitive requirements and consequences of existing assessments. The ultimate goal is to prescribe a set of specifications for selecting, creating, and scoring performance on items and tasks to support more valid and stable inferences about what students know and can do. Until we have such prescriptions, it will be impossible to ethically honor the principle of offering students a choice of varied ways to demonstrate their knowledge, as is being advocated by groups such as the National Council of Teachers of Mathematics (1993).

The most prevalent approach to analyzing the cognitive demands of particular assessment tasks involves having students talk about what they are doing and thinking either during or after the assessment (Baxter, Elder, & Glaser, 1995; Baxter et al., 1993; Hamilton, Nussbaum, & Snow, 1995). Webb (1995) has extended this approach to include analysis of student verbalizations as students interact in small groups while working on an assessment task. The original score distribution on an item should match the distribution of levels of understanding or reasoning assigned on the basis of the students' verbal protocols.

A less labor-intensive approach to identifying distinct cognitive components of performance being measured by particular items or sets of items is simply to examine the pattern of performance within and across sets of items. Factor analysis and/or logical analysis are used to identify the common attributes of tasks eliciting similar patterns of performance and the distinguishing attributes of tasks eliciting inconsistent patterns of performance. Snow et al. (1994) applied this approach to a set of mathematics and science items administered as part of the 1988 National Education Longitudinal Study.

This "patterns of performance" approach can be enhanced by the initial creation of sets of multiformat items that, in theory, should tap similar knowledge types or should make similar cognitive demands. Then, an examination of patterns of performance within and across sets of items will either support or contradict the hypothesized match between sets of items and cognitive components of performance. If patterns of performance do not support the hypothesized match between items and components, then the observed patters will provide an empirical basis for reformulating our theory about the cognitive demands of different types of items.

For this study, a construct-by-format matrix, suggested by Sugrue (1995), was used to generate a set of items targeting students' knowledge of the principle of Ohm's law. The matrix crosses three types of format (selection, generation, and explanation) with three types of knowledge (concepts, principles, and procedures), as well as metacognitive constructs and perceptions of self and task. This paper examines patterns of performance across selection, generation, and explanation items designed to measure the concepts of voltage, resistance, and current, as well as the principle that governs the relationships among them—Ohm's law. The context (written or hands-on) in which items were presented was also varied, and the influence of that context on performance was examined.

Methodology

Overview of the Larger Study

This paper is based on analysis of a subset of data gathered as part of a larger study designed to investigate the relative contribution of a number of cognitive constructs and test characteristics to variation in performance on a set of items and tasks designed to probe students' knowledge of Ohm's law. The cognitive constructs measured include knowledge of particular concepts, metacognitive variables, and perceptions of self and task. The test characteristics that were varied include format (selection, generation, and explanation), context of the assessment (written or hands-on), and working individually or in groups. The entire sample comprised 662 students in 7th and 8th grade in 21 classes in 5 schools in Los Angeles. Prior to assessment, students had three weeks of instruction on electric circuits. Teachers were told that the assessments would target students' knowledge of the concepts of voltage, resistance, and current, as well as the relationships among them, but they were not shown copies of the tests. The five teachers involved were given the freedom to teach the topic in their own way. Data were gathered on the instructional content and activities provided by each teacher and will be related to student performance in a future analysis.

At the end of the 3-week instructional period, each student completed two assessments: an individually-administered hands-on test consisting of two tasks and an individually-administered written test that contained a variety of items requiring students to select responses, generate responses, or justify their selected responses with written explanations. Half of the students took the written test first and half took the hands-on test first. One month later, the hands-on tasks were administered again to students working in groups of three, followed by another individual administration of the written test. A copy of the complete written test and a copy of the written instructions and follow-up written questions for the hands-on tasks are included in the Appendix.

Test Items Selected for This Analysis

In this paper, we consider data related to the following test items that probed students' knowledge of the concepts of voltage, resistance, and current:

1. items requiring students to select, from a set of pictured circuits, the circuit that exhibited the highest voltage, resistance or current (selection of circuits items); there were two such items for each concept (see items 2, 3, and 5, each with two sub-items, in the written test in the Appendix);

2. items requiring students to indicate whether the voltage, resistance, and current in a circuit would increase, decrease, or stay the same if different changes were made to the circuit (selection of prediction items); there were six such items for each concept (see item 6, with its 18 sub-items, in the written test in the Appendix);

3. items requiring students to draw diagrams of pairs of circuits that varied in their voltage, resistance, and current (generation of diagram items); there was one such item for each concept (see items 9, 10, and 11, in the written test in the Appendix);

4. extended hands-on tasks that required students to make pairs of circuits to match given specifications, using bags of materials provided. The materials in "Bag 1" were two 9-volt batteries, two 1.5-volt batteries, three bulbs in bulb holders, and seven wires with metal clips on the ends. The materials in "Bag 2" were two 9-volt batteries, two bulbs in bulb holders, three graphite resistors, and seven wires with clips on the end. Once students had constructed the real circuits, they were asked to draw diagrams of the circuits they made and answer three "selection" type items asking them to select which of the two circuits had the

highest voltage, resistance, and current; three "explanation" type items requiring them to justify the selections they had made regarding which circuit had the highest voltage, resistance, or current (see copies of the hands-on test instructions for "Bag 1" and "Bag 2" in the Appendix);

5. analogues of the hands-on tasks without the hands-on component (see items 13 and 14 in the written test the Appendix).

Scoring

Each selection type item was scored dichotomously. Diagrams not connected with the extended hands-on or analogue tasks were also coded dichotomously; to be coded correct, the circuits drawn in a diagram had to vary on the concept being measured: the voltage, resistance, or current had to be higher in one circuit. In the hands-on and written analogue extended tasks, the diagrams were scored on a 3-point scale. A score of 0 was assigned if a student did not draw two circuits that matched the goal of the task, (which was that one circuit should be brighter than the other); a score of 1 was assigned if the student drew circuits that matched the goal, but did not follow all of the instructions (for example, left out components or used components more than once); a score of 2 was assigned if the diagrams met the goal and did not violate any of the instructions.

Scores for written explanations related to the hands-on and analogue tasks were derived as follows. Explanations related to voltage were scored on a 4-point scale. A score of 0 was assigned if a student gave an irrelevant answer or displayed confusion over cause and effect such as "the voltage is higher because it is brighter." A score of 1 was assigned if the student mentioned batteries (but not the relative number) as the source of different voltage. A score of 2 was assigned if a student mentioned the relative number of batteries in each circuit. A score of 3 was assigned if a student mentioned the relative number of batteries and also referred to the relative power or voltage generated by the batteries.

Explanations related to resistance were scored on a 6-point scale. Again, 0 was assigned for irrelevant answers or answers confusing cause and effect. A score of 1 was assigned if the student referred to a difference in the number of "things" in the circuits. A score of 2 was assigned if a student was more specific and mentioned either wires or the distance the electrons had to travel. A score of 3 was assigned if the student mentioned graphite or bulbs as the cause of differences in resistance between the circuits. A score of 4 was assigned if the student described the number of items causing the difference in resistance. A score of 5 was assigned if a student referred to the exact numbers and types of items causing the difference in resistance.

Explanations related to current were coded along three dimensions (reference to voltage, reference to resistance, and reference to brightness) and a composite score based on a combination of these three codes was calculated. The voltage and resistance dimensions were coded on a 3-point scale where a score of 2 was assigned if a student referred to the more abstract concept of voltage or resistance, a score of 1 if the student referred to the more concrete concepts of batteries, bulbs or graphite, and a score of 0 if the student did not mention voltage or resistance (either abstract or concrete) as the cause of difference in the current in the two circuits. The third dimension coded (mention of brightness as a cause or effect of current) was coded dichotomously. The composite variable representing performance on current explanations was formed by summing the codes for the voltage and resistance dimensions, and giving half a point to a student who mentioned brightness but scored 0 on the voltage and resistance dimensions.

Scores on all diagram and explanation variables were converted to a 0 to 1 scale by dividing by the maximum possible number of points. This resulted in all score variables being on a 0 to 1 scale for analysis.

Samples for Analysis and Missing Data

A total of 519 students completed the written test, and 513 completed the hands-on test. If a student skipped an item within a test, a score of 0 was assigned.

Methods of Analysis

Short written test items. Mean *p*-values (percentage of students correctly answering each item) and correlations across subsets (three constructs and three formats) of the 27 written test items selected for analysis were examined. Patterns of *p*-values within types of items were also examined.

Extended hands-on and analogue tasks. First, the patterns of performance across components of the two hands-on tasks and the two analogue tasks from the written test were examined. The components of each of the four extended tasks were the diagrams drawn, the three selection questions (one relating to the voltage, one to the resistance, and one to the current in the constructed and drawn circuits), and the three written explanations (again, one relating to the voltage, one to the resistance and one to the current in the constructed and drawn circuits). Second, correlations among scores on different item types and constructs within and across versions (hands-on and analogue) were examined.

Results

Short Written Test Items

Pattern of performance across sets of items. Table 1 presents the mean *p*-values for the three types of short items on the written test for each construct being measured (voltage, resistance, and current). The difficulty of items varied depending on their format. The easiest item format was one that asks students to draw diagrams. The select circuit format was most difficult when voltage is the construct being measured; however, the select predictions format was easier for voltage than for resistance or current.

Table 2 presents the correlations among formats for each construct measured. Performance was most stable across the two selection types of items (r = .37 and .57), and was more stable across those two item types when the construct being measured is resistance (r = .57) than when the construct is voltage or current. Performance on items calling for the generation of diagrams was also more consistent with performance on selection items when the construct is resistance than when knowledge of voltage or current were being measured. Thus, we could say that students' knowledge of resistance was less sensitive to changes in test format than was their knowledge of voltage and current. Table 3 presents the correlations among knowledge constructs for each

Table 1

Mean *p*-Values on Selection and Diagram Items on Written Test (n = 519)

	Selection items		
Construct	Select circuit	Select predictions	Generate diagram items
Voltage	.33	.51	.75
Resistance	.47	.40	.49
Current	.40	.42	.65

Table 2

Correlations Among Short Item Formats for Each Construct (n = 519)

Correlation between pairs of formats	Construct measured		
	Voltage	Resistance	Current
Select circuit and select prediction	.37	.57	.37
Select circuit and generate diagram	.14	.35	.15
Select prediction and generate diagram	.24	.31	.17

Table 3

Correlations Among Constructs for Each Item Format (n = 519)

Correlation between pairs of constructs	Item format		
	Select circuit	Select prediction	Generate diagram
Voltage and resistance	.31	.52	.22
Voltage and current	.34	.50	.37
Resistance and current	.30	.47	.35

item format. Performance was most consistent across knowledge constructs when measured with the select prediction format (r = .47, .50, and .52).

Patterns of performance within sets of items. Table 4 presents the *p*-values for individual items of each selection type related to each knowledge component. The concept of voltage had the most variability in difficulty within the select circuit type of item. For all three knowledge constructs, there was considerable variation in performance within each set of six select prediction items. The largest discrepancy in performance was between two of the six select prediction items targeting the concept of voltage: only approximately 35% of students chose the correct prediction for three of these items, but almost 80% chose the correct prediction for two of the items. On closer examination, the pattern of performance across the 18 prediction items reflects particular strengths and gaps in students' knowledge of the three targeted concepts.

Table 4

P-Values for Selection Items on Written Test (n = 519)

Item number	Content/format	Item p-value	Mean p-value
2a	Voltage/select circuit	.25	
2b	Voltage/select circuit	.40	.33
3a	Resistance/select circuit	.44	
3b	Resistance/select circuit	.49	.47
5a	Current/select circuit	.43	
5b	Current/select circuit	.36	.40
6av	Voltage/select prediction	.34	
6bv	Voltage/select prediction	.78	
6cv	Voltage/select prediction	.36	
6dv	Voltage/select prediction	.77	.51
6ev	Voltage/select prediction	.48	
6fv	Voltage/select prediction	.36	
6ar	Resistance/select prediction	.51	
6br	Resistance/select prediction	.23	
6cr	Resistance/select prediction	.52	
6dr	Resistance/select prediction	.22	.40
6er	Resistance/select prediction	.39	
6fr	Resistance/select prediction	.55	
6ac	Current/select prediction	.36	
6bc	Current/select prediction	.61	
6cc	Current/select prediction	.39	
6dc	Current/select prediction	.56	.42
6ec	Current/select prediction	.37	
6fc	Current/select prediction	.23	

In the select prediction items, students were asked to predict the voltage, resistance, and current in a circuit if changes were made to the batteries, the bulbs, or if a glass rod was added to the circuit. Table 5 shows the mean p-values for prediction items involving changes to different circuit components. Students tended to do better when the component of the circuit being varied related most obviously to the attribute of the circuit to be predicted; for example, items where batteries were manipulated elicited a very high proportion of correct responses when voltage was the attribute to be predicted. The proportion of students correctly predicting the current in the circuit when batteries were manipulated was also relatively high, indicating perhaps that students had made a strong link between the concepts of voltage and current. Those same items (where batteries were manipulated) elicited the lowest level of performance when

Table 5

Mean p-Values for Select Prediction Items With Different Aspects of Circuits Varied (n = 519)

	Predict effects of . . .		
Construct	Change in batteries	Change in bulbs	Adding a glass rod
Voltage	.78	.35	.48
Resistance	.23	.53	.39
Current	.59	.37	.37

resistance was the attribute to be predicted, indicating that students have a general misconception of the link between batteries and resistance. When bulbs were the components of the circuit manipulated, student did better predicting the change in resistance in the circuit than the change in either voltage or current. This indicates that students were making an appropriate association between bulbs and resistance, but had a more fragile understanding of the connection between bulbs and voltage or current.

Summary. On short written test items, students did better on items where they had to draw diagrams depicting voltage and current than on items where they had to select from alternatives. For the concept of voltage, the select circuits format was the most difficult and the select prediction format was the easiest. Performance was most stable across item types for the concept of resistance. Performance was most stable across constructs when the select prediction format was used. However, composite scores across item within the select prediction format cloud the fact that performance was highly variable on items related to a particular concept. Within the select prediction format, students tended to do better when the component of the circuit being varied related most obviously to the attribute of the circuit to be predicted. This indicates that students' knowledge in this domain was not yet organized around more abstract concepts and principles.

Extended Hands-On and Analogue Tasks

We now turn to examine patterns of performance within and across extended hands-on tasks and analogue tasks (without the hands-on component).

Pattern of performance across components of the four extended tasks. Table 6 presents the mean scores on each of the components of the four extended tasks. The components are generation of diagrams, three selection questions based on the diagram (one related to voltage, one related to resistance, and one related to current), and three written explanation questions requiring justification of the choice made in the preceding selection question. Thus, there were seven scored components for each of the four extended tasks (two hands-on and two written analogues). The most notable pattern of performance here was the consistently lower scores on explanation items related to the concept of current than on explanation items related to voltage or resistance. In addition, scores on selection and explanation items related to voltage were lower for the second extended task than the first, whether in the hands-on or written analogue context. The second extended task was more difficult than the first, probably because it included graphite resistors in addition to batteries and bulbs and was therefore more novel. Performance on the selection items for Task 2 was higher in the hands-on version than in the written analogue, particularly for the concepts of resistance and current, indicating, perhaps, that for tasks with more novel surface features, the opportunity to work in a hands-on context aids performance.

Table 7 presents correlations among scores on the diagrams drawn in each of the four conditions. Correlations ranged from .27 to .41, the highest level of consistency being between diagrams drawn in the hands-on and analogue of Task 1 (r = .41), the lowest being between hands-on Task 1 and written Task 2 or hands-on and written Task 2.

Table 6

Mean Scores on Components of Extended Tasks, Phase 1 (n = 513)

	Written test						Hands-on test					
	Gen		Sel		Explan		Gen		Sel		Explan	
Construct measured	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Voltage			.60	.37	.52	.32			.55	.41	.45	.36
Resistance	.62	.68	.42	.45	.37	.42	.59	.71	.43	.56	.35	.45
Current			.47	.38	.25	.19			.52	.54	.24	.21

Note. T1, T2 = Task 1, Task 2; Gen = Generation of diagrams; Sel = selection question based on diagram; Explan = written explanation of choice made in selection question.

Table 7

Correlations Among Diagrams Drawn in the Four Task Conditions (n = 513)

	Hands-on Task 1	Hands-on Task 2	Written analogue Task 1
Hands-on Task 2	.34		
Written analogue Task 1	.41	.30	
Written analogue Task 2	.27	.27	.37

Tables 8, 9, and 10 present correlations among selection and explanation items within and across the hands-on or analogue conditions for each construct measured (voltage, resistance, and current). For voltage, correlations between selection and explanation items within any one test ranged from .58 (hands-on task 1) to .77 (hands-on task 2); correlations were higher for Task 2 regardless of hands-on or written analogue context (.74 and .77 for Task 2, compared to .58 and .44 for Task 1). For voltage, correlations across tasks were considerably lower than within tasks, ranging from .20 to .39. For resistance, within-task correlations were lower than for voltage, three of them being .56 or .57, and one (between the selection and explanation items in hands-on task 1) being only .44. Cross-task correlations were similar to those related to voltage. For current, both within- and cross-task correlations were considerably lower than for voltage or resistance. Within-task correlations between selection and explanation items for current ranged from .27 to .35. However, for current, the correlation between written analogue explanation items in Task 1 and 2 was higher (.51) than for voltage (.39) or resistance (.46).

Discussion

When the tests used in this study were originally designed, great care was taken to create a variety of types of items to target similar knowledge units. However, student performance was not stable across sets of short items originally conceived as calling for similar knowledge. While the content being targeted by a "select the circuit with the highest voltage" item, for example, might be the same as the content being targeted by a "select/predict what will happen to the voltage in the circuit if . . .", (both items require knowledge of voltage) the kind of reasoning a student has to do with the content is different

Table 8

Correlations Among Items Related to Voltage in Extended Tasks (n = 513)

	H1 Sel	H2 Sel	H1 Exp	H2 Exp	W1 Sel	W2 Sel	W1 Exp
H2 Sel	.39						
H1 Exp	**.58**	.31					
H2 Exp	.34	**.77**	.36				
W1 Sel	.25	.33	.23	.29			
W2 Sel	.30	.39	.26	.36	.38		
W1 Exp	.20	.26	.31	.26	**.44**	.34	
W2 Exp	.27	.32	.26	.35	.35	**.74**	.39

Note. H = Hands-on; W = Written analogue; Sel = Selection item; Exp = Explanation item.

Table 9

Correlations Among Items Related to Resistance in Extended Tasks (n = 513)

	H1 Sel	H2 Sel	H1 Exp	H2 Exp	W1 Sel	W2 Sel	W1 Exp
H2 Sel	.35						
H1 Exp	**.57**	.36					
H2 Exp	.25	**.40**	.33				
W1 Sel	.30	.32	**.23**	.26			
W2 Sel	**.22**	.31	.20	.20	.39		
W1 Exp	.26	.25	.39	.24	**.57**	.36	
W2 Exp	.19	.30	.29	.26	.27	**.56**	.46

Note. H = Hands-on; W = Written analogue; Sel = Selection item; Exp = Explanation item.

Table 10

Correlations Among Items Related to Current in Extended Tasks (n = 513)

	H1 Sel	H2 Sel	H1 Exp	H2 Exp	W1 Sel	W2 Sel	W1 Exp
H2 Sel	.20						
H1 Exp	**.34**	.14					
H2 Exp	.13	**.33**	.30				
W1 Sel	.16	.16	.13	.21			
W2 Sel	.09	.17	.07	.19	.26		
W1 Exp	.16	.08	.32	.29	**.27**	.23	
W2 Exp	.07	.00	.21	.27	.18	**.35**	.51

Note. H = Hands-on; W = Written analogue; Sel = Selection item; Exp = Explanation item.

enough to generate different score distributions for the two kinds of item. The most stable pattern of performance across short written item types was observed for the concept of resistance.

If performance is averaged across nonconverging sources of evidence, and if inferences about students' level of understanding of a particular concept are based on that "average" performance, then valuable information that would help inform subsequent instruction is lost. Of the short written test items examined in this study, the most useful diagnostic information was obtained from the pattern of performance across 18 "selection of prediction" items. That set of items measures the extent to which students associate changes in particular components of circuits with the voltage, resistance, and current in the circuit. Performance within this set of items revealed clear misconceptions about the links between batteries and resistance and between bulbs and voltage or current. A composite score on these 18 items, or even on the three groups of 6 items targeting voltage, resistance, and current, would have masked such misconceptions.

Examination of performance within and across the extended hands-on and analogue tasks indicates that, with the exception of selection items, the tasks were of equal difficulty regardless of opportunity to engage in hands-on activity. The opportunity to engage in hands-on activity made a difference in the more novel task, in terms of performance on selection items related to resistance and current. Consistency of performance on different item types was much higher within than across tasks, regardless of the context (written or hands-on). For example, performance on voltage selection items in the analogue tasks was highly correlated with performance on voltage explanation items in the analogue tasks; similarly, within the hands-on version of the tasks. However, performance on voltage items, even of the same format, across tasks (in either the written or hands-on context) was low. Within-task correlations on the extended tasks were highest for the concept of voltage and lowest for the concept of current. Thus, in both the short written test items and in the extended tasks, performance on the most complex concept (current) was generally the least stable.

More sophisticated analyses of these data, such as generalizability studies and confirmatory factor analysis, are planned. However, on the basis of the analysis of *p*-values and correlations presented here, one gets a sense of the sensitivity of performance to variation in format and context of test items. Not only does performance vary depending on how we ask students to display their knowledge, but the type and extent of variation in performance is not consistent across types or levels of knowledge. One implication of this study for test design is that before assuming the comparability of sets of items, one should examine patterns of performance within and across those sets of items.

The cognitive demands of items can vary even within formats targeting similar pieces of knowledge. However, performance across some item types, for example, multiple-choice and explanation items, may be more stable than we think. Evidence of such stability can justify decisions about which items to use when we need an efficient test for large-scale administration, or when we want the assessment to have higher face validity. In effect, examination of patterns of performance during the test design process can help promote the kind of validity that matters most, that is, construct validity.

References

Baker, E. L., O'Neil, Jr., H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist, 48*, 1210-1218.

Baxter, G. P., Elder, A. D., & Glaser, R. (1995). *Cognitive analysis of a science performance assessment*. (CSE Tech. Rep. No. 398). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Baxter, G. P., Glaser, R., & Raghavan, K. (1993). *Analysis of cognitive demand in selected alternative science assessments*. (CSE Tech. Rep. No. 382). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Bridgeman, B., & Rock, D. A. (1993). Relationships among multiple-choice and open-ended analytical questions. *Journal of Educational Measurement, 3*, 313-329.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Hamilton, L. S., Nussbaum, M., & Snow, R. E. (1995, April). *Alternative interview procedures for validating science assessments*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement, 31*, 234-250.

National Council of Teachers of Mathematics (1993). *Assessment standards for school mathematics: Working draft*. Reston, VA: NCTM.

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, *30*, pp. 215-232.

Snow, R. E. (1993). Construct validity and constructed-response tests. In R. Bennett & W. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 45-60). Hillsdale, NJ: Lawrence Erlbaum Associates.

Snow, R. E., Ennis. M., Kupermintz, H., & Talbert, J. (1994, April). *Opportunities to learn different components of mathematics achievement: Evidence from NELS:88.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement: Issues and Practice. 14*(3), 29-36.

Webb, N. M. (1995, April). *Using group collaboration as a window into students' cognitive processes*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Appendix

Hands-on Tests and Written Test

CRESST/UCLA ELECTRIC CIRCUITS HANDS-ON TEST, 1994

BAG 1

Name: ________________________________

Date: _____________________

1. Use the items in Bag 1 to make **two circuits** on the white paper mat on your desk.

Follow these rules:

- **Bulb A should be in Circuit A. Bulb B should be in Circuit B.**
- **Bulb A should be brighter than Bulb B**.
- **There should be one 9-volt battery in each circuit.**
- **Use all of the items in the bag but do not use any item more than once.** For example, if you put Bulb C in Circuit A, you cannot also put it in Circuit B.

2. In the boxes below, draw diagrams of the circuits you made.

Circuit A (brighter)	Circuit B (dimmer)

3. Why is Bulb A in Circuit A brighter than Bulb B in Circuit B? (Try to use scientific terms in your answer.)

____

____

____

4. Which of the two circuits you made has the **highest voltage**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME VOLTAGE**

Why? (Try to use scientific terms in your answer.)

____

____

____

5. Which of the two circuits you made has the **highest resistance**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME RESISTANCE**

Why? (Try to use scientific terms in your answer.)

____

____

____

6. Which of the two circuits you made has the **highest current**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME CURRENT**

Why? (Try to use scientific terms in your answer.)

____

____

____

CRESST/UCLA ELECTRIC CIRCUITS HANDS-ON TEST, 1994

BAG 2

Name: ______________________________

Date: ______________________

1. Use the items in Bag 2 to make **two circuits** on the white paper mat on your desk.

Follow these rules:

- **Bulb A should be in Circuit A. Bulb B should be in Circuit B.**
- **Bulb A should be dimmer than Bulb B**.
- **Use all of the items in the bag but do not use any item more than once.** For example, if you put one piece of graphite in circuit A, then you could not put three pieces in Circuit B (there would be only two pieces of graphite left to use).

2. In the boxes below, draw diagrams of the circuits you made.

Circuit A (dimmer)	Circuit B (brighter)

3. Why is Bulb A in Circuit A dimmer than Bulb B in Circuit B? (Try to use scientific terms in your answer.)

____

____

____

4. Which of the two circuits you made has the **highest voltage**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME VOLTAGE**

Why? (Try to use scientific terms in your answer.)

____

____

____

5. Which of the two circuits you made has the **highest resistance**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME RESISTANCE**

Why? (Try to use scientific terms in your answer.)

____

____

____

6. Which of the two circuits you made has the **highest current**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME CURRENT**

Why? (Try to use scientific terms in your answer.)

____

____

____

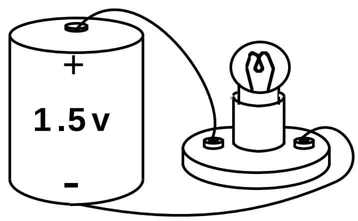
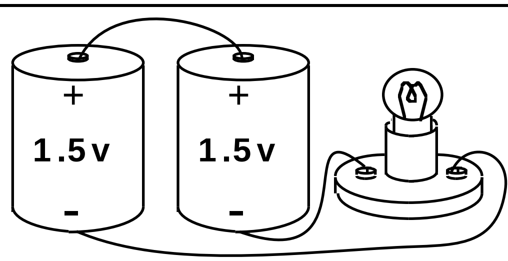
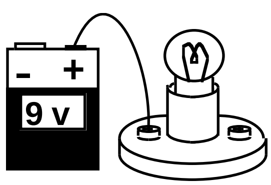
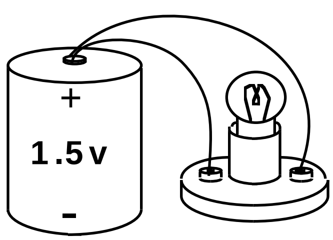
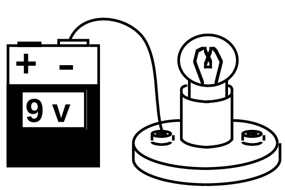
Form A

CRESST/UCLA ELECTRIC CIRCUITS WRITTEN TEST, 1994

Name: ______________________________

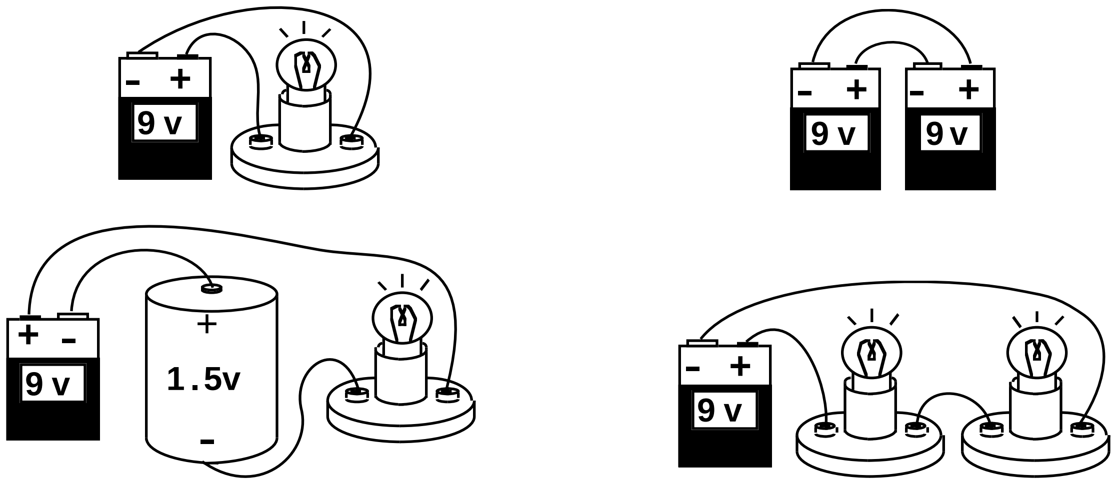
Date: ____________________

1. For each of the following five circuit diagrams, circle **YES** or **NO** to indicate if it is a **complete circuit,** and circle **YES** or **NO** to indicate if the bulb will light.

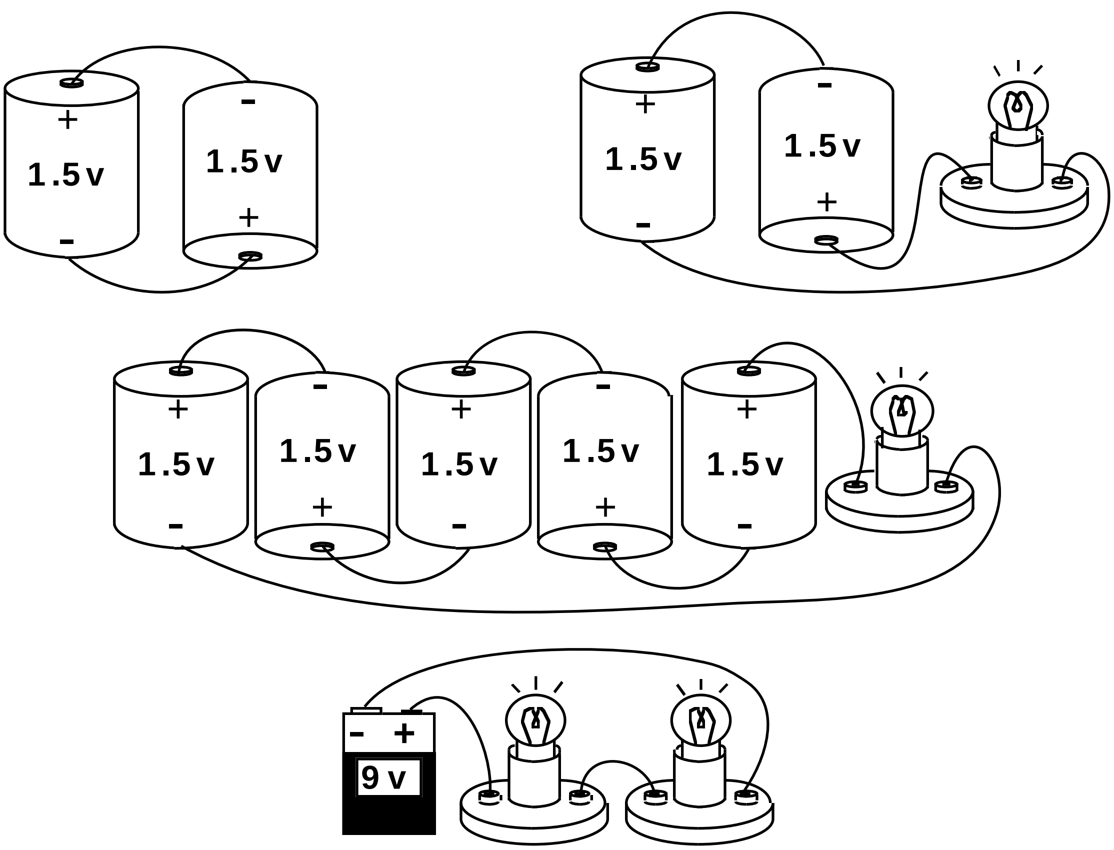
	Complete circuit? **YES** **NO**	Will the bulb light? **YES** **NO**
	Complete circuit? **YES** **NO**	Will the bulb light? **YES** **NO**
	Complete circuit? **YES** **NO**	Will the bulb light? **YES** **NO**
	Complete circuit? **YES** **NO**	Will the bulb light? **YES** **NO**
	Complete circuit? **YES** **NO**	Will the bulb light? **YES** **NO**

2. For each of the following two sets of circuits, (a) and (b), circle the circuit that has the **highest voltage**. Assume that all circuits are properly connected.

(a) Circle the circuit with the highest **voltage**:



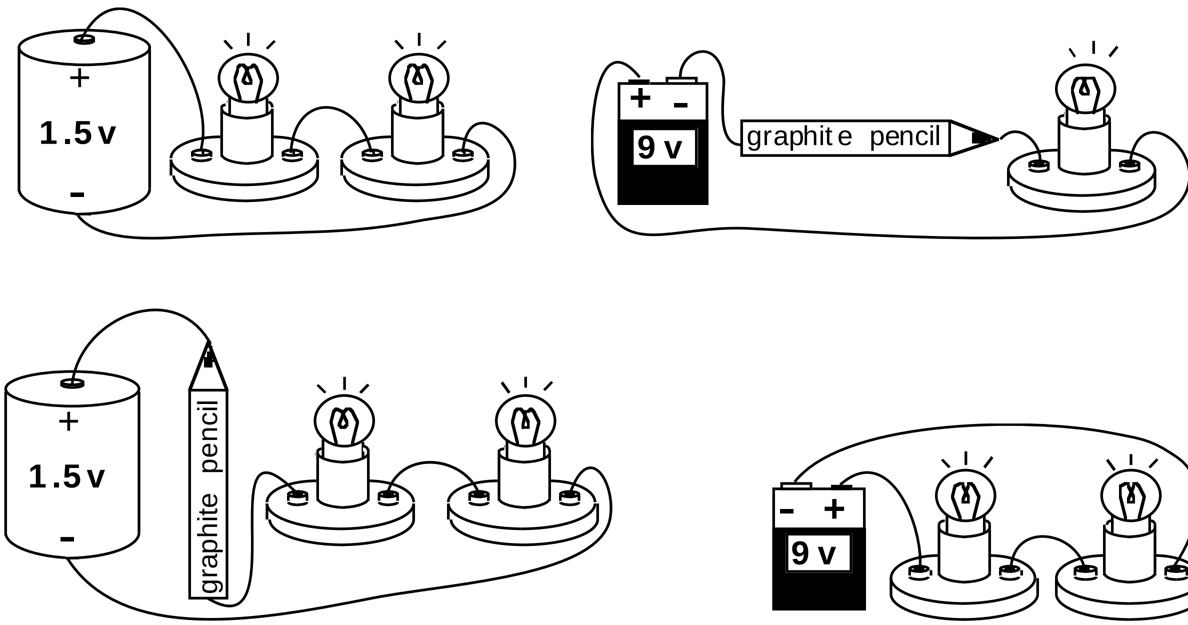
(b) Circle the circuit with the highest **voltage**:



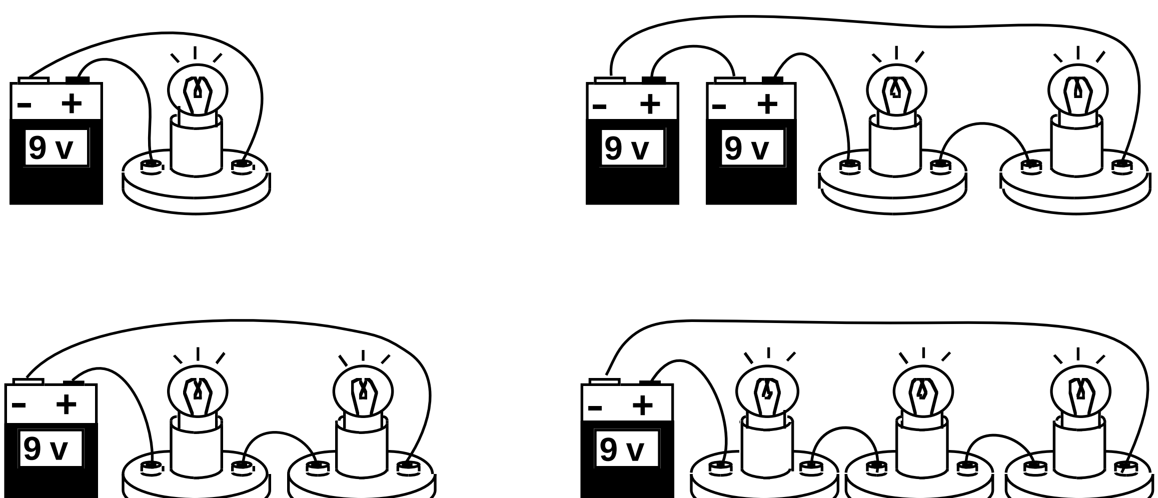
---

3. For each of the following two sets of circuits, (a) and (b), circle the circuit that has the **highest resistance**. Assume that all circuits are properly connected.

(a) Circle the circuit with the highest **resistance**:

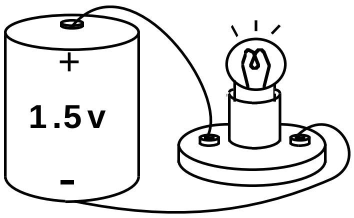


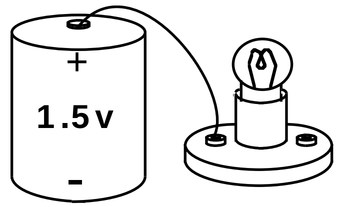
(b) Circle the circuit with the highest **resistance**:

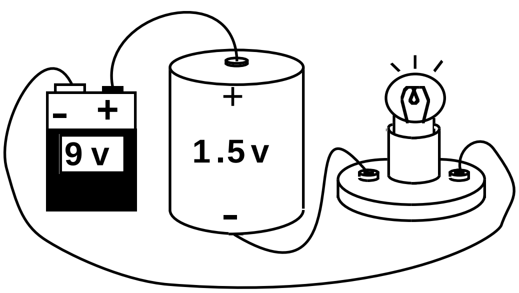


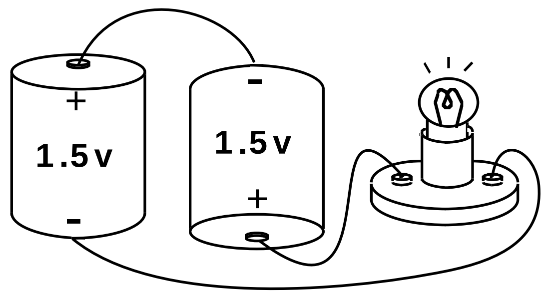
---

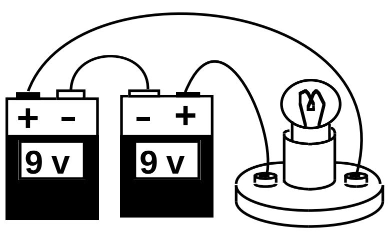
4. Circle any of the following seven circuits that have **current** flowing through them. You may circle **more than one** circuit.

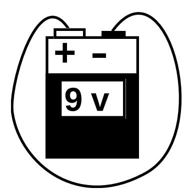


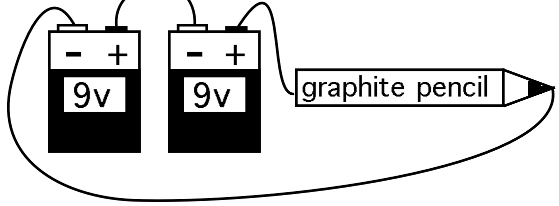






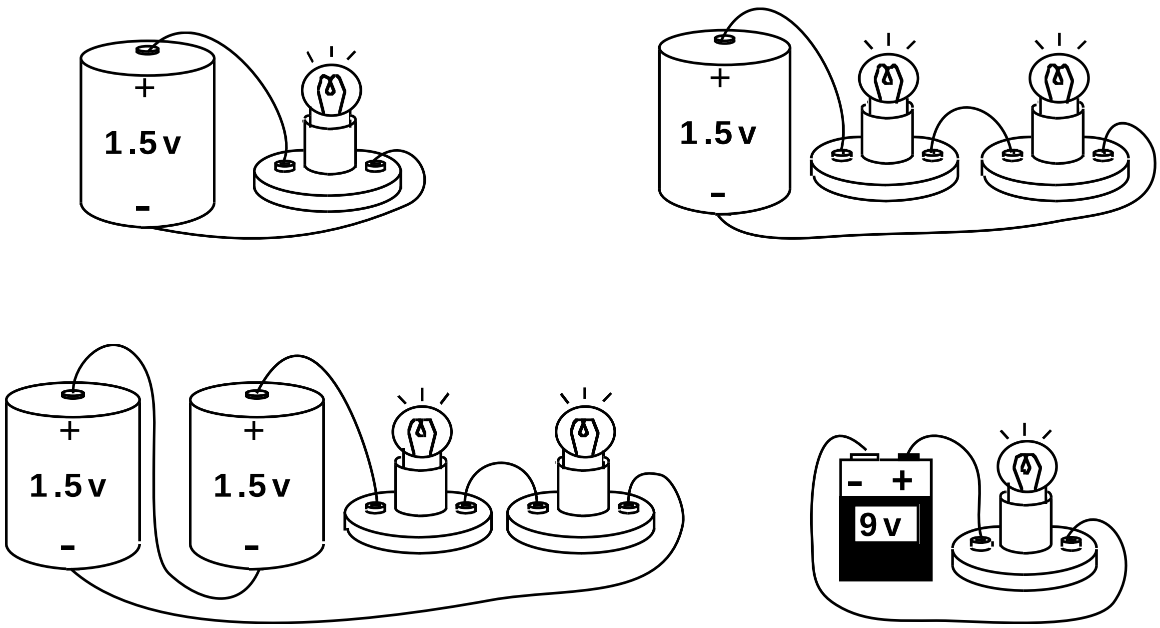




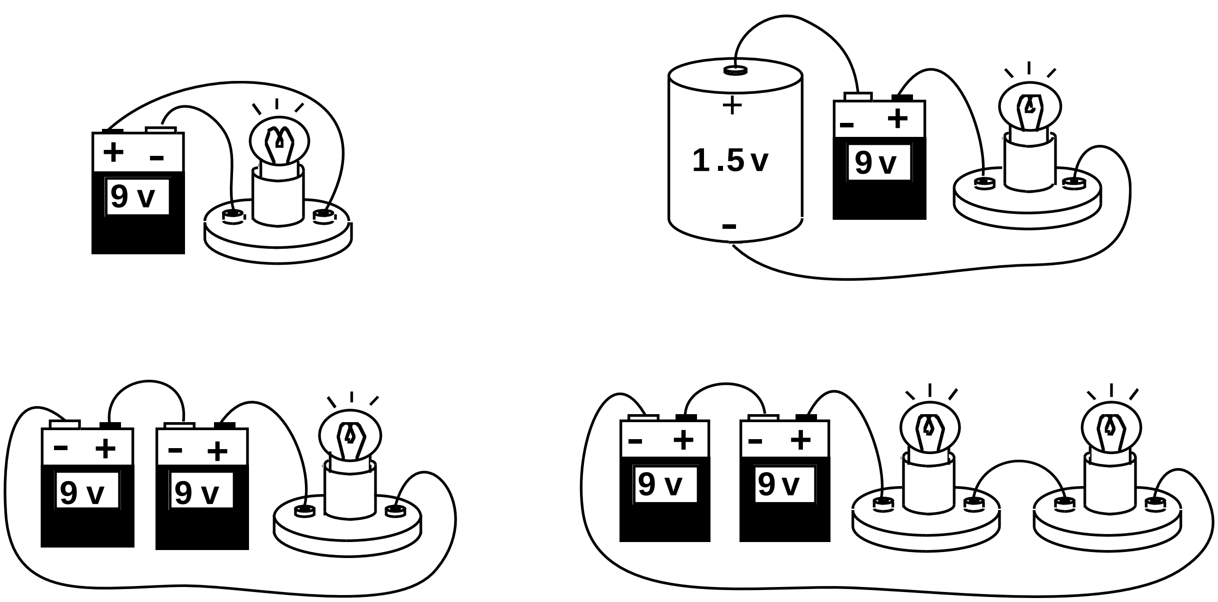


5. For each of the following two sets of circuits, (a) and (b), circle the circuit that has the **highest current**. Assume that all circuits are properly connected.

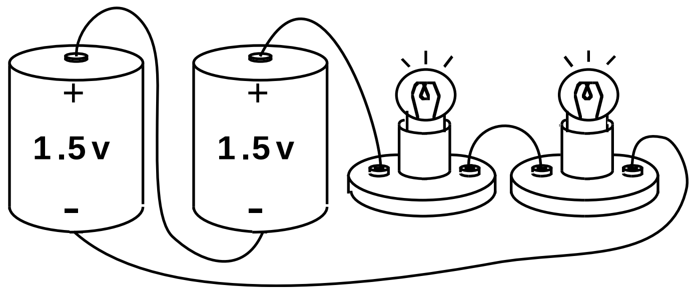
(a) Circle the circuit with the highest **current**:



(b) Circle the circuit with the highest **current**:

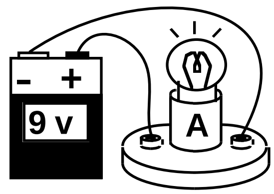


6. Predict what will happen to the voltage, resistance, and current in the following circuit if each of the changes listed in the chart is made. Circle **INCREASE, DECREASE,** or **NO CHANGE,** in each box in the chart. Assume that the circuit is properly reconnected after a change is made.



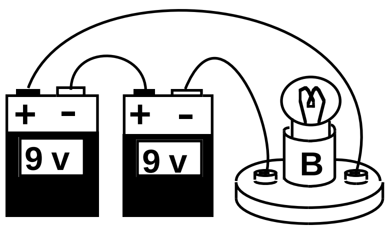
What will happen if you	Voltage	Resistance	Current
add another bulb?	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE
add a 9-volt battery?	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE
remove one bulb?	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE
remove one battery?	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE
add a glass rod?	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE
remove both bulbs?	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE	INCREASE DECREASE NO CHANGE

7.



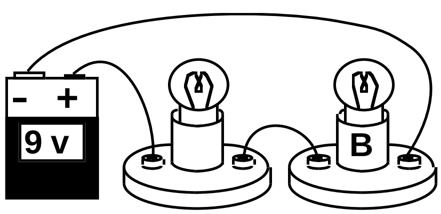
In the circuit above, Bulb A is **bright.**

For each of the following four circuits, indicate whether Bulb B will be **brighter** than Bulb A, **dimmer** than Bulb A, the **same** as Bulb A, or **will not light** at all. Circle the correct answer and then give a reason for your choice. Try to use scientific terms in your explanations.



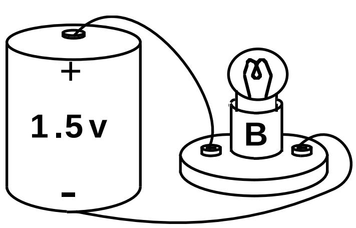
BRIGHTER **SAME** **DIMMER** **WILL NOT LIGHT**

Why?



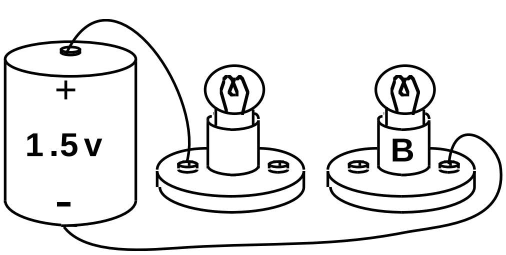
BRIGHTER **SAME** **DIMMER** **WILL NOT LIGHT**

Why?



BRIGHTER **SAME** **DIMMER** **WILL NOT LIGHT**

Why?



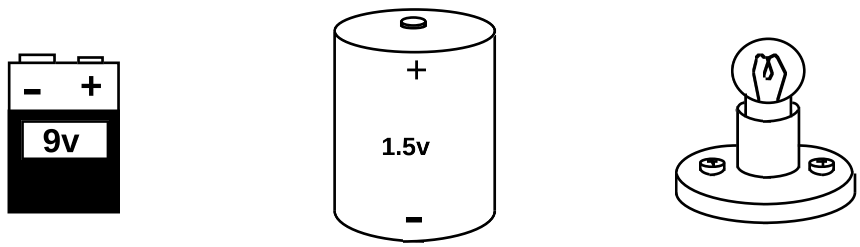
BRIGHTER **SAME** **DIMMER** **WILL NOT LIGHT**

Why?

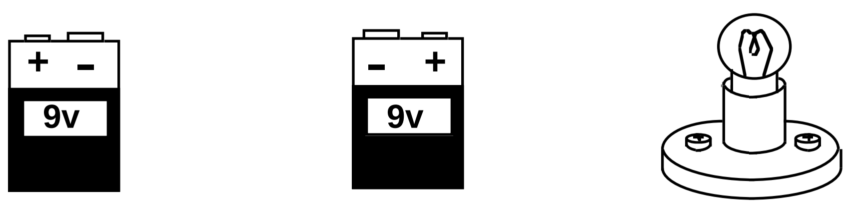
---

8. For each of the following two sets of batteries and bulbs, (a) and (b), **draw wires** to connect all of the batteries and the bulb to make a **circuit** that will **light the bulb**.

(a)



(b)



9. Draw two circuits in the boxes labeled Circuit 1 and Circuit 2 below.

- **Circuit 1 should have higher voltage than Circuit 2.**
- **Each circuit should have at least one battery and one bulb.**

Circuit 1 **(higher voltage)**	**Circuit 2** **(lower voltage)**

10. Draw two circuits in the boxes labeled Circuit 1 and Circuit 2 below.

- **Circuit 1 should have higher resistance than Circuit 2.**
- **Each circuit should have at least one battery and one bulb.**

Circuit 1 **(higher resistance)**	**Circuit 2** **(lower resistance)**

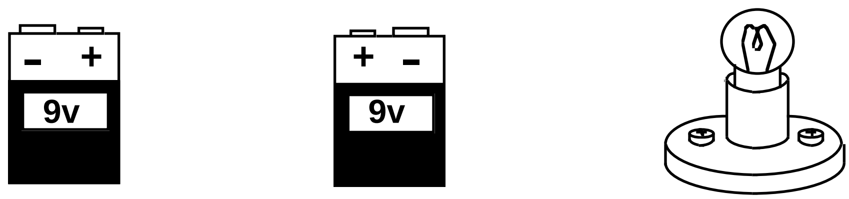
11. Draw two circuits in the boxes labeled Circuit 1 and Circuit 2 below.

- **Circuit 1 should have higher current than Circuit 2.**
- **Each circuit should have at least one battery and one bulb.**

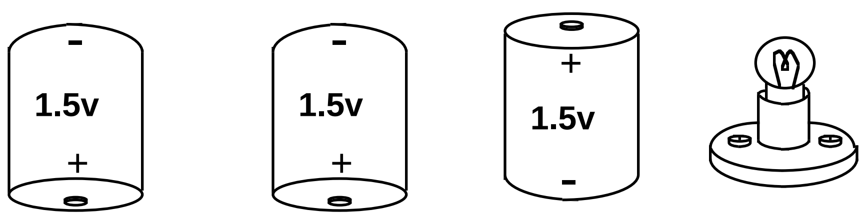
Circuit 1 **(higher current)**	**Circuit 2** **(lower current)**

12. For each of the following two sets of batteries and bulbs, (a) and (b), **draw wires** so that **current** flows through the circuit. Then **draw arrows** on the wires to indicate the **direction in which the current is flowing**.

(a)



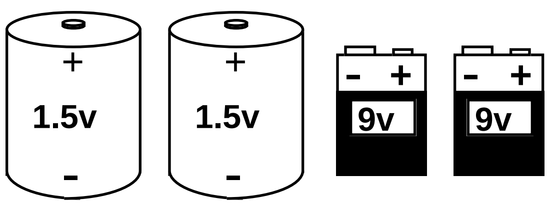
(b)



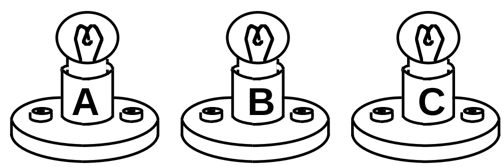
13. (a) Use the items drawn below (batteries and bulbs) to draw **two circuits** in the boxes labeled Circuit A and Circuit B. **Follow these rules:**

- **Bulb A should be in Circuit A. Bulb B should be in Circuit B.**
- **Bulb A should be brighter than Bulb B**.
- **There should be one 9-volt battery in each circuit.**
- **You must draw the wires needed to connect up the items in each circuit.**
- **Use all of the items but do not use any item more than once.** For example, if you put Bulb C in Circuit A, you cannot also put it in Circuit B.

Batteries



Bulbs



Draw the circuits in these boxes:

Circuit A (brighter)	**Circuit B (dimmer)**

13. (b) Why will Bulb A in Circuit A be brighter than Bulb B in Circuit B? (Try to use scientific terms in your answer.)

____

____

____

13. (c) Which of the two circuits you drew has the **highest voltage**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME VOLTAGE**

Why? (Try to use scientific terms in your answer.)

____

____

____

13. (d) Which of the two circuits you drew has the **highest resistance**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME RESISTANCE**

Why? (Try to use scientific terms in your answer.)

____

____

____

13. (e) Which of the two circuits you drew has the **highest current**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME CURRENT**

Why? (Try to use scientific terms in your answer.)

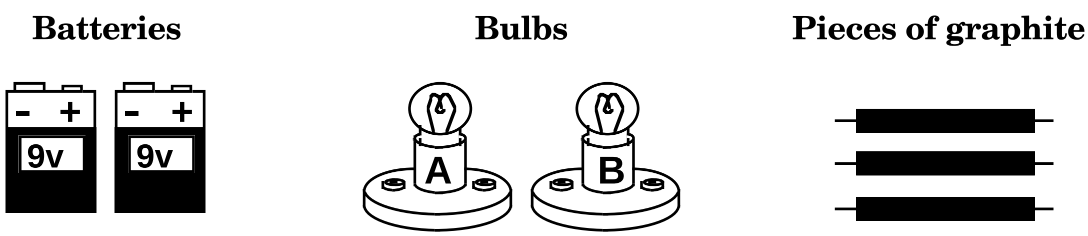
____

____

____

14. (a) Use the items drawn below (batteries, bulbs, and pieces of graphite like the "lead" in your pencil) to draw **two circuits** in the boxes labeled Circuit A and Circuit B. **Follow these rules:**

- **Bulb A should be in Circuit A. Bulb B should be in Circuit B.**
- **Bulb A should be dimmer than Bulb B**.
- **You must draw the wires needed to connect up the items in each circuit.**
- **Use all of the items but do not use any item more than once.** For example, if you put one piece of graphite in circuit A, then you could not put three pieces in Circuit B (there would be only two pieces of graphite left to use).



Draw the circuits in these boxes:

Circuit A (dimmer)	Circuit B (brighter)

14. (b) Why will Bulb A in Circuit A be dimmer than Bulb B in Circuit B? (Try to use scientific terms in your answer.)

14. (c) Which of the two circuits you drew has the **highest voltage**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME VOLTAGE**

Why? (Try to use scientific terms in your answer.)

____

____

____

14. (d) Which of the two circuits you drew has the **highest resistance**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME RESISTANCE**

Why? (Try to use scientific terms in your answer.)

____

____

____

14. (e) Which of the two circuits you drew has the **highest current**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME CURRENT**

Why? (Try to use scientific terms in your answer.)

____

____

____