The Development of Constructed Response Astronomy Assessment Items

Matthew M Steele, John Merrill, Kevin Haudek, and Mark Urban-Lurain

1 Introduction

Concept inventories comprised of multiple choice questions are widely used in physics and astronomy to measure students pre-instruction knowledge and post instruction gains (Hestenes, Wells, & Swackhamer, 1992; Maloney, OKuma, Hieggelke, & Van Heuvelen, 2001, and others). Concept inventories are valuable because they address big ideas in the discipline with which students often struggle. The multiple choice construction of these instruments, however imposes structural limitations on the number and nature of misconceptions and mixed conceptual models which may be presented as item distractors (Smith & Tanner, 2010). Constructed response questions provide one means of addressing these limitations.

In the effort to develop constructed response questions in the field of astronomy, we have adapted items from well-known concept inventories. These items are comprised of two, three-question sets. By adapting items from existing concept inventories the resulting items are targeted as a supplement to the original instrument, providing an additional perspective with a deeper view on student thinking of the relevant items.

In this work we discuss the selection and adaption multiple choice items as part of the new question design of the AACR Question Development Cycle (appropriate reference), the testing of the resulting items, and a summary of patterns in student thinking that appeared in the body of collected responses.

2 Study Design and Procedure

For this work we focus on a set of questions drawn from the Light and Spectroscopy Concept Inventory (Bardar, Prather, Brecher, & Slater, 2007, hereafter LSCI). The selected items probe student thinking on connections between emission and absorption line features and the physical properties of their source objects. A classical test theory analysis by Schlingman, Prather, Wallace, Rudolph, and Brissenden (2012) revealed a number of items on the LSCI which displayed high post-instruction difficulty values and low post-instruction discrimination values. For these items students had difficulty answering the question after instruction and their success in doing did not correlate strongly with their score on the entire LSCI. For this study we have selected three high difficulty/low discrimination for adaptation to constructed response items, in order to investigate why students struggle with concepts targeted in the corresponding items.

2.1 Item selection and Adaptation

The items selected for adaptation are LSCI items 2, 17, and 21. The revisions to the text of these items when adapting from multiple choice to constructed response format were left as minimal as possible. Item 2 was adapted by simply removing the provided answer and distractors. Item 17 was reworded to remove the "which of the following format language. For item 21 a prompt for the "what kind of object" was replaced with query for "what physical properties". The original multiple choice answers provided object descriptions as permutations of the temperature (hot/cool) and density (dense/diffuse). With this construction simply removing the provided answers could allow the students to answer with a category of object (star/nebula/galaxy) rather than focus on the intended object properties. The constructed response version of the question was rephrased to explicitly focus on the source object's physical properties.

The exact formats of both the original LSCI items and the revised constructed response items may be found in appendix A. Hereafter in this work the constructed response items will be referred to by names that describe the key concept underlying the item, rather than the LSCI item numbers. The constructed response item derived from LSCI item 2 is referred to as the "object color and absorption features" item. The item produced from LSCI item 17 is known as the "emission line wavelength comparison" item. Finally question adapted from LSCI item 21 is called the "emission line sources" item. The items are discussed in the order which they were administered for this study, rather than the LSCI arrangement.

2.2 Item Administration

To test the converted constructed response items, they were administered to in two non-major university astronomy courses. On section was small (N$\sim$ 50) on-line summer session, the other a medium enrollment (N$\sim$ 150) traditional classroom course during the regular academic year. In the online section questions were given both pre- and post-instruction in order to solicit the greatest diversity of student responses. The items given only post-instruction in the tradition section. Both sections saw the items as part of regular weekly homework assignments. In total the response set for the three items contain 43 pre-instruction and $110 - 113$ post-instruction responses. The resulting responses are analyzed for their textual content in the following section.

3 Results and Analysis

This section describes the the exploratory rubric development procedure and the analysis of student responses of the three items. The rubric development and scoring processes is discussed in subsection 3.1. Each the results of scored student responses to each constructed response in considered individually in the subsequent subsections.

3.1 Rubric Development and Scoring

Student responses to constructed response items can display a wide range of ideas representative of both focused content understanding an loosely collected conceptual associations. The development of exploratory rubric were assembled with the goal of capturing as much student thinking on the items content as possible without excluding or preferring different levels of sophistication. As such the bins in the rubrics are intentionally fine grained, with the goal of coding for the presence of small coherent ideas in the student response. Bins in the rubric are non-mutually exclusive so a single student response may be coded in multiple bins. Under this scheme complex thought will be indicated by the presence of rubric bins which maybe connected to form a chain or reasoning. For example a complete well-constructed response may be coded for a rubric bin that indicates a comparison, another bin that indicates a reason for the comparison. A less coherent student response may then be coded as a collection of bins which correspond with an astrophysical object's physical properties.

The first step in constructing a exploratory rubric is examine a body of student responses for each constructed response item. The first response in the item set is examined and a rubric bin is defined for each explicit coherent concept it contains. The threshold for coherence is that reader developing the rubric is able to identify and describe the concept with a brief "handful of words" definition. To be viable the concept must be readily identifiable in isolation from the remainder the the response text. Since the rubrics used in study will ultimately be used by a automated machine learning system, ambiguous or implied concepts will not prompt a bin be added to the rubric, since the automated scoring system does not have the ability to interpret implied information. The set of responses is progressed through adding bins necessary as new concepts are encountered. If during the initial assembly of the rubric it becomes apparent that a concept is being used in fundamentally different ways indifferent responses the initial bin maybe redefined to a narrower concept and additional bins added for the new concept use case.

Once the first pass through the response set has been completed the initial rubric is examined to check for bins sufficiently similar that they cannot be reliably distinguished using individual student responses as test cases. Any such similar bins are consolidated. Then the rubric bins are sorted into broad categories such as "comparisons", "reasoning", or "physical properties" dependent on the content and/or use of the member bins. The specific type and functions of these categories will vary from item to item depending on the content and structure of the question. It should be noted that these categories are not explicitly coded for in student responses, they are merely used as a logistic aid for the coding and analysis procedures. The exploratory rubrics can be large, so having a way to organize the bins can speed the scoring process and provide an initial framework for the analysis. The rubrics developed for the three constructed response items considered in this work are presented in appendix B.

The completed rubric is then used to score the student response set which was employed in its development. As in the rubric construction phase a concept must be coherent, identifiable in isolation, and not require inference beyond the what the text makes explicit to be scored for a particular rubric bin. The results of the scoring of the test set are given along with exploratory analysis is sections 3.2, 3.3, and 3.3.

The rubrics developed in this work are exploratory in nature. The goal in their construction is both to simultaneously evaluate the constructed response question and for a preliminary picture of student thinking around the specific system the items address. While they are designed to be identify fine grained concepts and minimize the affects of coder interpretation and inference they have not been validated with in depth student interview data nor tested for reliability with multiple coders. As such the results they produce should taken as informative of the range of concepts contained within students responses, but not a robust probe of student understanding of the items content.

3.2 Emission Line Sources

When a bound electron transitions from high energy excitation state one with lower energy state, the atom must release energy equivalent to the difference in the two excitation states. If the atom exists in relative isolation, the most common means of releasing this energy is in the form of a photon. Since the energy levels available to an electron in a atom are quantized, each transition between two given level will result in the production of a photon with identical energy. To the observer a photons energy is measurable through its wavelength. A collection of photons produced by a specific excitation state transition from an element will all have the same wavelength. When viewed through device like a prism or diffraction grating, a collection of photons with identical wavelength appears as a bright line of a specific color. This type of electromagnetic radiation is there for known as emission line radiation, and the object that produces it is an emission line source. In order for an astrophysical systems to produce emission line radiation two conditions must be met; 1) the object must be hot enough for electrons in the objects atoms to exist in an initial high energy state, 2) the object must be diffuse enough such that photon emission can serve as a major mechanism for de-excitation and not be dominated by atom-atom collisional de-excitation. The emission line source item, both in its original LSCI and revised constructed response version is targeted at prompting the students to identify these two conditions.

The rubric for the emission line source constructed response item, adopted from LSCI item 21, is comprised of 15 bins (see table 2). These bins may

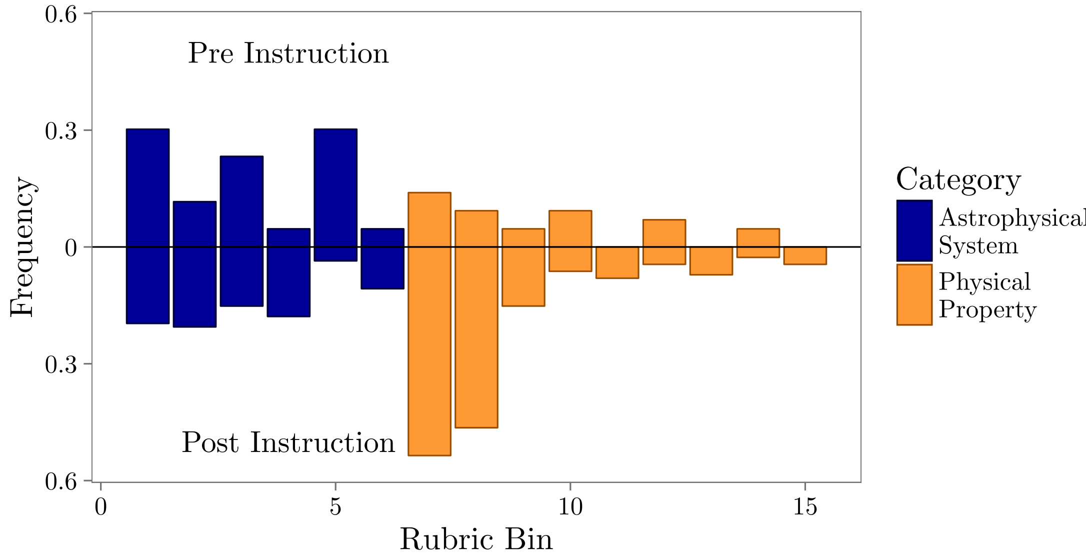

Figure 1: Frequency of responses for the emission line sources exploratory rubric. Pre-instruction responses are displayed in the upper portion with the post-instruction frequencies in the lower portion. The numeric rubric bin IDs can be found along with descriptions in table 2.

roughly be divided into two categories; rubric bins which address the nature and interactions of the astrophysical system, and bins that discuss the physical properties of the source. For illustration the “object type” bin which notes a discussion of whether the emission is produced by a nebula, star, planet or other class of object falls in the the astrophysical system category, while the “density” bin covering mentions of the density of the source media is of the physical property type.

The original LSCI multiple choice answers are covered by rubric bin 7 “temperature” and bin 9 “density”. It should be noted that neither the “temperature” bin, or the “density” distinguishes the quantity of the relevant measure in the rubric scheme. A response that discusses ”high temperature” and one that describes “low temperature” will each be coded only for the “temperature” bin. Therefore it is impossible for the exploratory rubric presented here to exactly reproduce the original LSCI item answers. The objective of the exploratory rubric is to test if students identify “temperature” and “density” as the key physical properties which determine if a source will produce emission lines.

The results of the test set of student responses is summarized in figure 1. The reference ID numbers for each bin in the figure are provided in table 2. The upper portion of the figure displays the frequency of rubric bin occurrence in

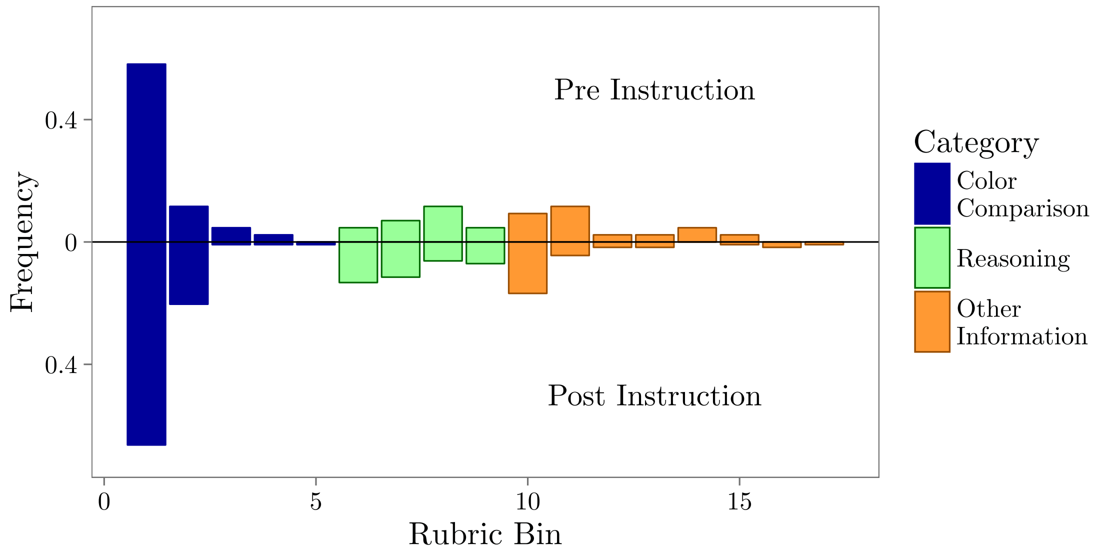

Figure 2: Frequency of responses for the object color and absorption features exploratory rubric. Pre-instruction responses are displayed in the upper portion with the post-instruction frequencies in the lower portion. The numeric rubric bin IDs can be found along with descriptions in table 3.

the pre-instruction response set, while the lower portion gives the rubric bin frequencies for the post-instruction responses.

The two rubric bins which correspond to the original LSCI answers, "temperature" and "density" have frequencies of 43% (14% pre, 54% post) and 12% (5% pre, 15% post) respectively. Only 8% (0% pre, 12% post)of student responses were scored for both the temperature and density rubric bins. After temperature the second most frequently occuring bin was "chemical composition" which occured in 36% (9% pre, 46% post) of student responses.

Responses from the pre-instruction set appeared to be disproportionately focused on concepts grouped in the "astrophysical system" category of bins. Specifically the "light/matter" (30% pre, 20% post), "properties of light" (23% pre, 15% post), and "energy content" (30% pre, 4% post), were over represented.

3.3 Object Color and Absorption Features

Hot dense objects like stars produce light through thermal radiation. The energy released in the core of the star gradually migrates to a layer of the star known as the photosphere through a combination of thermal conduction, convection, and photon scattering processes. By the time the energy leaves the

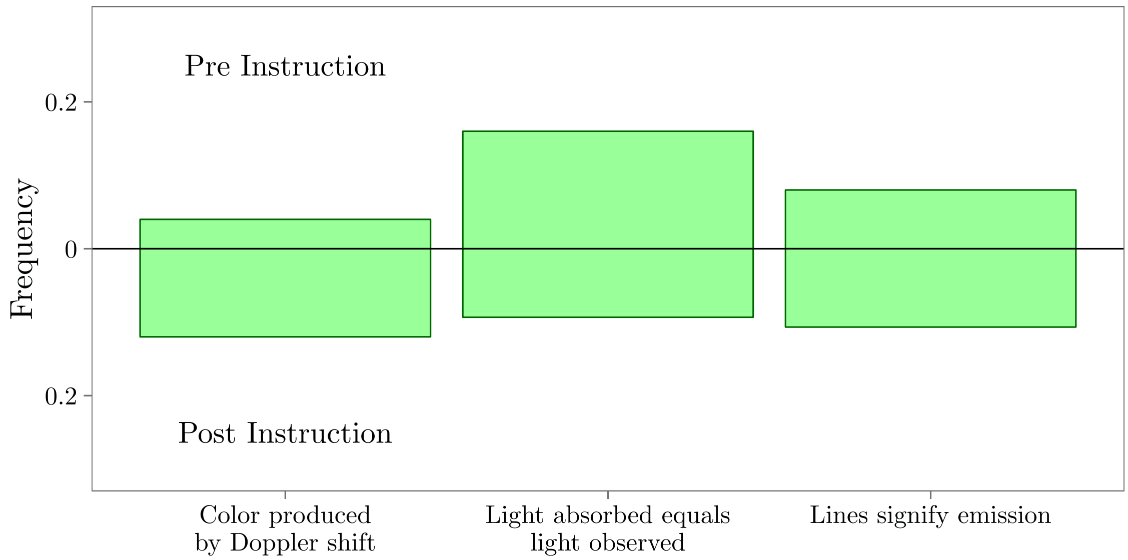

Figure 3: Frequency of the reasoning category rubric bins for responses also scored with the "x is blue, z is red" bin. The frequencies represent the fraction of response within the "x is blue, z is red" subset that were scored for the give rubric bin. Pre-instruction responses are displayed in the upper portion with the post-instruction frequencies in the lower portion.

photosphere in the form of photons, thermodynamic processes have shaped their energy density function to a characteristic distribution known as a thermal, or blackbody, distribution. The emissions that correspond with this distribution are known alternatively as blackbody radiation, thermal, or continuum emission. (The term continuum is used to distinguish the continuous emission over a wide range of wavelengths in opposition to the discrete specific emission of emission line radiation.) The observed both the peak wavelength and luminosity of any object whose emissions display a thermal spectrum is set by the temperature of the location from which the photons originate. In the case of a star this is the photosphere. A number of processes may add additional features to the spectrum of a star, for our purposes the most important is absorption by gases in the star's atmosphere. The absorption process is the inverse to emission line mechanism described in the previous section; bound electrons in low energy state absorb photons that have the specific discrete energies necessary to excite the electron to a higher energy state in its atom. This process selectively removes specific wavelengths from the stars spectrum, as photons with the corresponding energies are absorbed by the stars atmosphere and hence to do not make it to an outside observer. Since these removed wavelength appear as dark lines when the stars light is dispersed into a spectrum they are known as absorption lines. To an observer the color of a star is set by the wavelengths of light at which the star is most luminous. These dominant wavelengths are set by the stars thermal continuum emission. The absorption line processes are a secondary effect which reprocess a portion of the stars emissions, but do not directly set peak emission wavelengths.

In the object color and absorption line features item students are tasked with recognizing that thermal continuum emission is the primary driver determining the color of a star. The spectra the students are presenting with in the question are images of dispersal spectra which the students will be familiar with, but which do not contain sufficient information to determine the peak wavelengths of the stars light, and therefore its color. In both the LSCI version and the constructed response adaptation the students attention is focused on “the dark absorption line spectra”, with intention the student recognize the the stars color may not be identified by absorption line features alone.

The 17 bins that make up the rubric for the “object color and absorption features” item based on LSCI item 2, may divided into three categories. The first of these categories is a color comparison of the two stars considered in the item. An example bin from this category is “x is blue, z is red”. Responses in this bin compare the two stars and determine star x is bluer than star z. The second category of rubric bins is reasoning bins. Reasoning bins describe the rationale for identifying the color comparison. The final category of bins is “other information” indicating the response discusses concepts not directly relevant to the central color comparison.

In this exploratory rubric the multiple choice answers of the original LSCI item map directly to rubric bins 1,2,4 & 5. As seen in the scoring summary in figure 2, rubric bin 1, “x is blue, z is red”, is by far the most frequently occurring bin being scored in 64% (58% pre, 66% post) of the test response set. The opposite comparison in bin 2, “x is red, z is blue” occurs in 18% (12% pre, 20% post)of responses. Especially noteworthy is that ”not enough information” bin appears in less that one percent of all student responses.

Figure 3 displays the frequency of reasoning category rubric bins among responses that were also coded for “x is blue, z is red”. The combined pre/post frequency of each of the three reasoning bins is approximately equal (“Color produce by Doppler shift” 10%, “Light absorbed equal light observed” 11%, “Lines signify emission” 10%). Responses from the post instruction set preferentially describe the stars color as being a product of Doppler shift (4% pre, 12% post), while the pre instruction responses favor an explanation involve the color of light absorbed by an object being equivalent to the color observed (16% pre, 9% post). In neither group, however, do a majority of the students provide a reason for selecting “x is blue, z is red” (28% pre, 32%) post.

3.4 Emission Line Wavelength Comparisons

The specific quantized energy levels available to a bound electron in an atom are set by the composition of the atoms nucleus (the number of protons, and to a

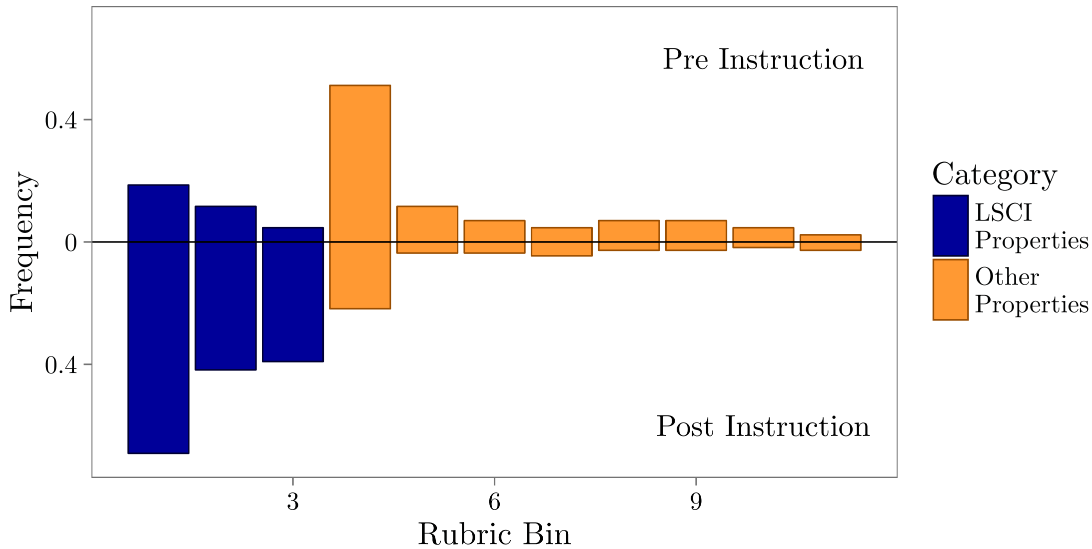

Figure 4: Frequency of responses for the emission line wavelength comparisons exploratory rubric. Pre-instruction responses are displayed in the upper portion with the post-instruction frequencies in the lower portion. The numeric rubric bin IDs can be found along with descriptions in table 4.

lesser extent the number of neutrons) and the number of bound electrons. Since each element has a distinct number of protons, it will also have a unique set of quantized energy levels available to it's electrons for any given ionization state. It follows that a unique set of energy levels will produce a unique set of differences between levels, and consequently will produce a unique set of photon wavelengths when electrons transition between levels. To the observer an ionization state of a given element will produce a characteristic emission line spectra which can be used to identify the presence of elements in distant astrophysical sources. It is important to note, though, that the absolute wavelength of an emission line may be modified by a Doppler shift of the emission. When there is relative motion between an emission source and the observer the observed wavelength of a photon will be shifted dependent on the magnitude and direction of the motion. When the motion is decreasing the source-observer distance the observed wavelength is higher than its rest frame wavelength (bluer in color, known as blueshift), when the motion increases the distance the observed wavelength is longer (redder in color, known as redshift). While Doppler shift is able to change absolute wavelength of a photon it does not change the relative wavelengths of photons from the same source. This means that observing a set of emission lines that display wavelength relative to one another it is not only possible to measure

Table 1: Permutations of composition, temperature, and motion rubric bins

bin permutation	**pre (%)**	**post (%)**	**all (%)**
Composition, Temperature, & Motion	0	15	11
Only Composition & Temperature	7	14	12
Only Composition & Motion	2	15	12
Only Temperature & Motion	2	3	3
Only Composition	9	26	25
Only Temperature	2	10	8
Only Motion	0	5	4
Permutation Total	23	87	69

what element produced the collection of lines, but of how the source is moving relative to the observer.

The temperature of a gas comprised of a specific element determines the ionization states of the its constituent atoms. Therefore the collection of emission lines and identifiable ions of an element will vary with temperature. At any given temperature, however, multiple ionization states may exist, so a determination of the presence of an ion or even collection of ions, is insufficient to make a measure of temperature. In order to determine the temperature a ratio of ions for a given element must be made. This ion ratio is only measurable through measurement of the luminosity of emission lines, not there wavelength.

Both the LSCI and constructed response versions of the emission line wavelength comparison question ask the student to consider the physical properties of an emission line nebula which are directly measurable from the emission lines wavelengths alone. The LSCI version provides the three parameters considered above (composition, motion, and temperature) while the constructed response version requires students to determine the parameters relevant for consideration themselves.

The exploratory rubric for the ”emission line wavelength comparisons” item contains 11 bins which for convenience may be divided into two categories; “LSCI properties” and “Other properties”. The bins in each category describe physical properties of the emission line source. Those assigned to the “LSCI properties” category describe physical properties which were present in the original LSCI multiple choice answers (see appendix A.3). The “other properties” items discuss source object parameters not considered in the LSCI item.

The original LSCI responses were permutations of three properties; composition, temperature, and motion. In the rubric presented here each of these three properties is assigned it’s own bin (bins 1,2, & 3 respectively). Figure 4 shows that these bins are the most frequently occurring in the post-instruction sample (composition – 69%, temperature – 42%, motion – 39 %) and all display significantly higher representation than in the pre-instruction sample (composition – 19%, temperature – 12%, motion – 5%). The “other property” bin with the highest frequency is that of “color”, occurring in 51% of pre-instruction responses and 22% of the post-instruction set.

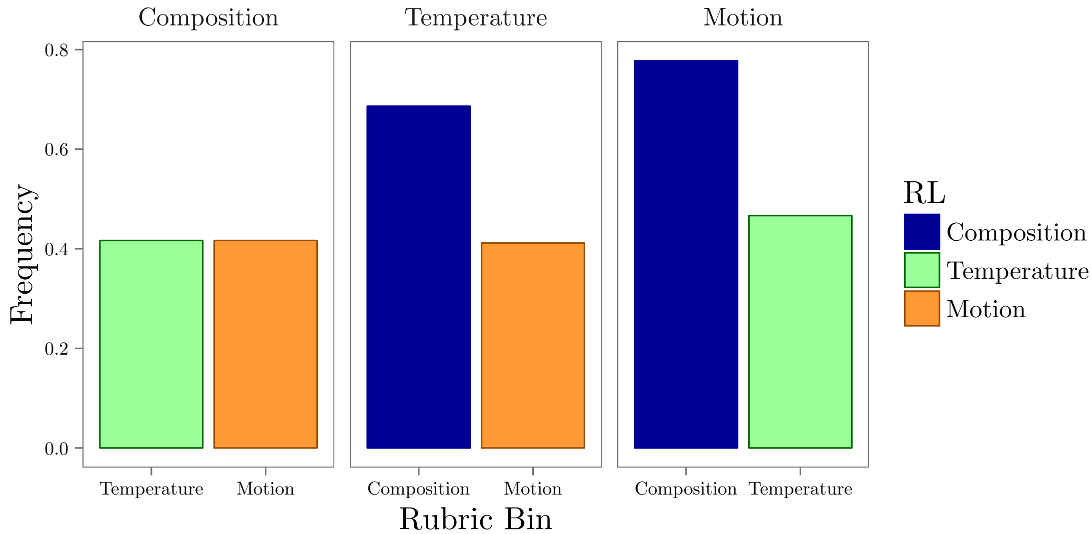

Figure 5: Conditional frequency of responses for the emission line wavelength comparisons exploratory rubric. All responses in a given panel are positively scored for the titled rubric bin. Conditional Frequency = $N(\text{Bin A}|\text{Bin B})/N(\text{Bin B})$. For which N is the number of responses, "Bin A" is the rubric bin listed on the horizontal axis, and "Bin B" the the panel title rubric bin.

To directly compare the exploratory rubric results with expected results for the original LSCI answers the co-occurrence the composition, temperature, and motion bins must be examined. The results of this exploration are summarized in table 1. Of note is that 87% of the post instructional responses include a permutation of these three rubric bins, while only 23% of pre-instructional response do. It should also be noted that permutations not covered by the the original LSCI answers account for 14% of student responses (9% pre, 16% post).

Another tact of examining the co-occurrence of the composition, temperature, and motion bins is to examine if the mention of one physical property makes it more likely that either of the other two is also mentioned. As such we define the conditional frequency as the number of responses scored for a rubric bin given the scoring of conditional rubric bin normalized to the number of responses scored with the conditional bin (Conditional Frequency = $N(\text{Bin A}|\text{Bin B})/N(\text{Bin B})$). The results of this comparison are given in figure 5. Student responses that are coded for composition are equally likely to mention either temperature or motion, while the majority of students who mention temperature or motion will also mention composition.

4 Discussion and Conclusion

The student responses to the emission line source item present here fail to correspond to the possible answers presented by LSCI multiple choice version. As noted in section 3.2 only 8% of student responses included the two concepts, "temperature" and "density" permutations of which comprised all possible options of the LSCI text. The data considered in the present study is unable to distinguish if this discrepancy may be attributable to the change of language of the constructed response version. Since the constructed response language is more specific, inquiring about "physical properties" rather than "type of object", it seems difficult to explain how it would prompt a less focused set of student responses.

Independent of the effects of the language change the results of this study do call into question the students ability to interpret the text of the constructed response emission line source item. The diversity of concepts suggest that some portion of the students may have interpreted the item as asking "what can we learn/measure about the physical properties of the object" rather than the intended "what physical conditions are necessary to produce emission lines". This line of reasoning is further supported by the "composition" rubric bin being the second most frequently occurring bin. While composition is measurable physical property of the emission line source, it does not serve as condition for production of emission lines. (Assuming the composition the students refer to is more specific than normal, non-dark matter.)

The pre-instruction focus on "astrophysical object" rubric bins on the emission line source is likely do to instructional effects. In the courses used in this study students begin the term studying the properties of light and basic light/matter interactions. The most common pre-instructional rubric bins were "light/matter interaction", "energy content", and "properties of light", all three of which are less prevalent in the post-instructional set. As such the over-representation of the "astrophysical object" bins in the pre-instruction set may simple be a case of may students writing what they knew rather than directly responding to question.

In order to test the possibility of students misinterpreting the language of the item to inquire about measurable physical parameters rather than required conditions for emission line production, the question should be modified for future work. A refined question could read: "If the light coming from a distant object produces a bright emission line, what do we know about the physical properties of the distant object which are necessary to produce this emission?" With such a wording it should be possible to investigate whether the question wording or the underlying concepts are the source of the observed student difficulties.

For the object color and absorption features item students do not appear to struggle with the interpretation of the question. In the combined pre/post-instruction sample 84% of students provide a color comparison covered by the option in the original LSCI item. Rather the difficulty lies in providing a correct answer, with less than one percent of the post-instruction students provide a response coded in the "not enough information" bin. It is entirely possible that in the constructed response format students do not view “not enough information” as an acceptable answer.

If the assumption is made that students view there possible choices in the constructed response as being limited to responses which would be coded as alternatively “star x is blue, star z is red”, “star x is red, star z is blue”, and or “stars same color”, the set of responses is problematic. Given a scenario where the thermal continuum emission of the two stars in negligible for the color comparison, perhaps the stars share identical thermal emission and so the secondary affects of the the absorption features are important in color comparison, students still select the incorrect “x is blue, z is red” over “x is red, z is blue” by better than a 3 to 1 margin. The student responses do not provide a clear reason for the so overwhelmingly selecting the “x is blue, z is red” comparison. The 32% of post instructional students who describe this comparison in their responses and provide reasoning to support it are split along three possible explanations.

Even if the premise is accepted that the current structure of the constructed response version of the object color and absorption feature question is flawed, the most probable conclusion to draw from the set of test responses is that students have significant difficulty with the mechanisms and effects of line absorption processes.

To examine the issues seen here in the object color and absorption features item two modifications to the question are necessary. First the question should be modified with language explicitly allowing the permissibility of a “not enough information response”. A phrasing such as “Is it possible to determine the color of the stars from these spectra? If so, what can you determine about the colors of the two stars?” As a second modification the question should include a prompt for an explanation of reasoning. If the result of student struggles with absorption features presented in the work is robust, more complete descriptions of student reasoning will be necessary to determine the origin of the difficulties.

For post-instructional students the composition, temperature, and motion rubric bins in the exploratory rubric for the “emission line wavelength comparison” item are sufficient to capture the majority of student thought. The added permutations available in the constructed response format boost the response set coverage of 71% for a LSCI-like set of combinations to the 87% in the response set presented here. From these results it is likely that the LSCI format form of the question is largely representative of student thinking on the topic, with the constructed response version provide more complete coverage without the prompting affect potential to multiple choice items.

If the students’ interpretation of the question and ability to discuss the relavent concepts for the emission line wavelength comparison item is promising, the frequency of correct responses is not. Only 14% of the post-instruction response correctly included a discussion of composition and motion, but not temperature. This difficulty appears to stem from the fact that both temperature and motion have similar frequencies in the post-instruction sample ($\sim$ 40%). Further students who discuss one of these two concepts are equally likely to discuss the other as is the general response population. If this were a multiple choice item such statistics would be suggestive of guessing (and may still be for constructed response answers). It seems likely however that the students in the test sample were able to identify the key measurable parameters, but struggled with the specifics as of the emission line wavelength comparison.

The pre-instruction responses for the wavelength comparison item on the other hand do not give strong indications of student thinking. The most popular rubric bin is “color”, which may be a result of question administration order and the implied association of color and spectral line features from the color and absorption line features item. Beyond that possible feature the pre-instruction responses appear to be a grab bag of physical properties possibly indicative that student to not have well formed ideas of what may be learn from the comparison of emission line features to laboratory references.

In the next phase of this program the three items will be considered for alteration, re-administered, and predictive rubrics developed for inclusion in the AACR automated scoring system. The “emission line sources” item will be modified to specify a focus on the physical parameters necessary for emission line production, rather than any measurable (or inferable) property. The “object color and absorption features” will be tested with split versions alternately containing and omitting a focusing statement on the thermal continuum. Both versions of this item will also include a prompt for the reasoning supporting the comparison. The “emission line wavelength comparison” item is ready for additional data collection and predictive rubric development in it’s current form.

The three items adapted from the LSCI presented here all display promise to provide additional insight into student thinking on astronomical EM spectrum when presented in an constructed response format. Two of the items have shown significant problems in post-instruction student thinking, and the third suggests an incomplete picture of measurable properties. On the completion of future work all three items should provide a valuable supplement to the LSCI for instructors to assess there students’ learning.

Author Note

This work is supported by the National Science Foundation (DUE 1323162 and 1347740). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

Bardar, E. M., Prather, E. E., Brecher, K., & Slater, T. F. (2007). Development and validation of the light and spectroscopy concept inventory. *Astronomy Education Review*, *5*(2), 103–113.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141–158.

Maloney, D. P., OKuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69*(S1), S12–S23.

Schlingman, W. M., Prather, E. E., Wallace, C. S., Rudolph, A. L., & Brissenden, G. (2012). A classical test theory analysis of the light and spectroscopy concept inventory national study data set. *Astronomy Education Review*, *11*, 010107.

Smith, J. I., & Tanner, K. (2010). The problem of revealing how students think: concept inventories and beyond. *CBE-Life Sciences Education*, *9*(1), 1–5.

A LSCI Items and Constructed Response Forms

A.1 Item: Emission Line Sources

Adopted from LSCI Item 21.

> **Original:** If the light coming from a distant object produces a bright line emission spectrum, what kind of object is it?
>
> a Hot and dense.
>
> b Cool and dense.
>
> c Hot and diffuse.
>
> d Cool and diffuse.

> **Constructed Response:** If the light coming from a distant object produces a bright emission line, what do we know about the physical properties of the distant object?

A.2 Item: Object Color and Absorption Features

Figure 6: LSCI Item 2 (Bardar et al., 2007)

Both the original version of LSCI Item 2 and the constructed response adaptation refer to figure 6.

> **Original:** Consider the dark line absorption spectra shown below for Star X and Star Z. What can you determine about the colors of the two stars? *Assume that the left end of each spectrum corresponds to shorter wavelengths (blue light) and that the right end of each spectrum corresponds with longer wavelengths (red light).*
>
> a Star X would appear blue and Star Z would appear red.
>
> b Star X would appear red and Star Z would appear blue.
>
> c Both stars would appear the same color.
>
> d The colors of the stars cannot be determined from this information.

Constructed Response: Consider the dark line absorption spectra shown below for Star X and Star Z. What can you determine about the colors of the two stars? Assume that the left end of each spectrum corresponds to shorter wavelengths (blue light) and that the right end of each spectrum corresponds with longer wavelengths (red light).

A.3 Item: Emission Line Wavelength Comparisons

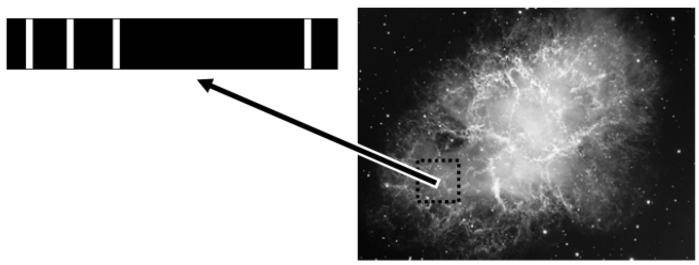

Figure 7: LSCI Item 17 (Bardar et al., 2007)

Both the original version of LSCI item 17 and the constructed response adaptation presented here refer to figure 7

Original: The bright line emission spectrum shown above is characteristic of the region of the nebula marked in the drawing. By comparing the *positions* of the lines in the spectrum to a known laboratory spectrum on Earth, which of the following properties of the nebula can be *directly* determined?

a Motion towards or away from Earth only.

b Temperature only.

c Chemical composition (type of atoms) only.

d Motion and chemical composition.

e Motion, temperature, and chemical composition.

Constructed Response: The bright line emission spectrum shown above is characteristic of the region of the nebula marked in the drawing. By comparing the positions of the lines in the spectrum to a known laboratory spectrum on Earth, what properties of the nebula can be directly determined?

B Constructed Response Exploratory Rubrics

Table 2: LSCI item 21 constructed response exploratory rubric

ID	Rubric Bin	Description
1	light/matter interaction	Response includes a description of a process of light/matter interaction; eg. emission, absorption.
2	object type	Emission lines identify astrophysical object class; eg.stars, nebulae.
3	properties of light	Emission lines provide information on properties of light; eg. wavelength, luminosity.
4	state of matter	Emission lines signify a the sources state of matter; eg. gas, solid.
5	energy content	The emission lines provide information on the amount and/or source of the objects internal energy.
6	secondary object	The emission lines provide information on secondary sources in the system; eg. cool gas clouds, planets.
7	temperature	The emission lines provide a measure of the objects temperature.
8	chemical composition	The emission line allow for a determination of the objects chemical composition.
9	density	The emission line provide a measure of the source's density.
10	distance	The emission lines allow for a measure of the distance between the source and the observer.
11	mass	The emission lines provide for a determination of the systems mass.
12	geometry	The emission lines provide information pertaining the the source's size, shape, etc.
13	kinematics	The emission lines provide a measure of the objects kinematics.
14	age	The emission lines allow for a determination of the source's age.
15	color	The emission lines provide information on the object's color.

Table 3: LSCI Item 2 constructed response exploratory rubric

ID	Rubric Bin	Description
1	x is blue, z is red	Star x has a blue color, star z has a red color.
2	X is red, z is blue	Star x has a red color, star z has a blue color.
3	stars different colors	The too stars have different colors, but how they differ is not specified.
4	stars same color	That stars are the same color.
5	not enough information	The two given spectra do not provide sufficient information to determine the stars' color.
6	color produced by Doppler shift	The color observed is produced by a Doppler shift of the stars light; redshift, blueshift.
7	light absorbed is not observed	The star's light appears the colors of light which are not absorbed by its atmosphere.
8	light absorbed equals light observed	An observer sees the stars as being the same color as the light it absorbs.
9	lines signify emission	The dark lines in spectra correspond to the wavelengths of light which the star is emitting.
10	color indicates temperature	The star's color provides a measure of its temperature.
11	color indicates luminosity	The star's color provides a measure of its luminosity.
12	color indicates object type	The star's color is determined by it's object class; eg. blue giant, red dwarf.
13	color indicates distance	The star's color provides a measure of its distance from the observer.
14	color indicates age	The star's color corresponds to the stars age; eg. blue stars are young, red stars are old.
15	color indicates lifespan	The stars color provides and indication of its lifespan.
16	color indicates mass	The stars color can be used to determine its mass.
17	color produced by thermal emission	The color of the star is produced by thermal radiation, not absorption line features.

Table 4: LSCI item 17 constructed response exploratory rubric

ID	Rubric Bin	Description
1	composition	A comparison of observed wavelength to a known reference allows for the identification of the source objects chemical composition.
2	temperature	The comparison provides a measurement of temperature.
3	motion	The comparison provides a measure of the object kinematics.
4	color	The comparison allows a measure of the objects color.
5	luminosity	The comparison provides a measure of the source's luminosity.
6	distance	The comparison provides a measure of the source's distance.
7	density	The comparison provides a measure of the source's density.
8	geometry	The comparison provides information on the objects size, shape, etc.
9	object type	The comparison allows a determinination if the object is a star, planet, nebula, etc.
10	age	The comparison provides a measure of the source's age.
11	mass	The comparison provides a measure of the source's mass