# *A New Method for Assessing Student Knowledge*

Stephen H. Levy
Saint Peter's University
Jersey City, NJ, USA
drslevy@gmail.com

Abstract: We often say that one person knows more than another about a given topic -- whether the character of an acquaintance, the news headlines, or academic subjects like math, literature, chemistry, or history. These claims, though largely true, are imprecise because 1) they may lack clear evidence, 2) certain concepts are not at hand, or 3) the attempts to measure the evidence, if available, are not made. In this paper, we address these issues. The central idea is of a concept of set of traits in one's mind. Different traits distinguish different concepts. Interpreted as what two people know, the two concepts comprise the different amounts of knowledge of the one person (the student) and the other (the expert), or the same person at different times. With various levels of attention, we observe the concept traits of others and ourselves. Formally and informally, we see what others say or not say, write or not write, reveal or not reveal. To promote this method, we define several metrics, such as concept differences, concept distance, knowledge ratio, and ignorance ratio. Examples are drawn from math, statistics, computing, movies, song lyrics, among other subjects. The method complements other methods of assessment.

Keywords: *concept traits, concept differences, concept distances, knowledge, knowledge ratio, measurement, assessment*

**A New Method for Assessing Student Knowledge**

We often say that one person knows more than another about a question, and yet a third person knows still more. We may be thinking of our knowledge of everyday matters like the character of an acquaintance, highlights in the news, or topics in academic subjects like math, literature, chemistry, or history. In education, we optimistically believe that the student knows more at the end of the course than she or he did in the beginning.

These claims, though usually true, are typically vague and imprecise. We reach them by informal and formal observation of others' speech, conversations, and writing. In this paper, we develop a method to make these claims somewhat more precise -- but, of course, exactitude is an elusive goal. The method develops intuitive ideas and draws on ideas from set theory, philosophy, computing, and education.

Central to the method is the idea of a concept. A concept is somewhat analogous to a set or class in set theory. (The logician Gottlob Frege, in fact, defined a set as the objects that fall under a concept (begriff).) But there is a difference. *Whereas a set is a collection of objects of any kind however conceived, a concept is a collection of ideas of traits or characteristics*. These traits may seem to go together "naturally" (as a physical description of a person -- tall, muscular, and brown-eyed) or they may simply be conceived together in a person's mind (as the physical description of a person in a specific place -- the beach -- at a given time.) The traits may or may not be a definition of a term. Just as a set can in turn have other sets as its members, a concept can in turn have other concepts as its constituents.

A concept trait (characteristic or attribute) differs from a question item. A question item can cite or suggest traits -- "What is the average salary of CEOs of large cap corporations?" -- and, conversely, mentioning a trait can suggest a question item -- "The average salary … is...", but they are not the same. The number of traits we know or can think of (whether expert or student) generally far exceeds the number of questions we can formulate in speech or writing. This points to the common observation that as helpful as question items are, they do not provide the only means of assessing student, class, or expert progress in creative and critical thinking, knowledge, or knowledge acquisition.

For our purposes, I take *knowledge* to be an agreement in traits between anyone's thought and the thought of a recognized expert in the subject. This common conception of knowledge sidesteps the philosophic issue of whether knowledge and truth are about objective features of the world or subjective features of our minds. The thought of an expert serves as a standard of knowledge regardless of which view is correct. Admittedly, this standard itself changes to some extent from one expert to another and within the same expert over time.

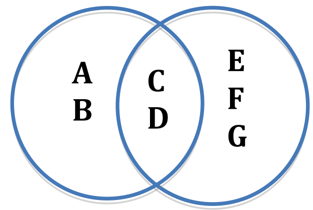
The statistics developed in this paper are designed to assist us in estimating how these traits in our thinking reflect our knowledge in various fields and on assorted topics. Developing the statistical measures may aid in developing question items, but that is not their chief purpose.

Given two concepts, they may or may not share some traits in common. But there are two kinds of *differences* between them. There are traits in the first concept not in the second, and traits in the second concept not in the first. Consider the concepts *range* and *median*. Both share the traits "set of values," "lowest value," "highest value," but the range introduces the one trait "their difference," whereas the median introduces the two traits "arranged in order" and "middle value." Neither concept shares the additions of the other. The *distance* between the two concepts is the sum of the number or count of differences of both kinds. The distance between range and median is 1 + 2 = 3. The distance can be conceived as the number of steps in a mental transformation from either concept to the other by removals or additions of traits.

**Figure 1 Distance = Sum of Counts of Differences**

**= Sum of Removals & Additions of Difference**
**To Transform One into Other**

**Before A B C D**
**After C D E F G Distance = 5**




Now change the two concepts. The first is the student's somewhat erroneous concept of an idea, the second is the expert's correct concept. The distance between the concepts now signifies the degree of the student's error, or the distance from the expert's concept or what the expert *knows*. But this distance or error comprises two types -- what the student mistakenly *adds* to the concept shared, and what the student mistakenly *ignores* from the expert's concept.

Again, consider the *median.* The expert has the correct concept above, whereas the student's concept contains the shared traits -- "set of values," "lowest value," and "highest value." but omits the other two traits of median. Rather, perhaps thinking of spread and mean, the student adds the traits "their difference" and "divide by two." Applied to 2, 3, 5, 10, and 20, the expert takes the median to be 5, while the student takes it to be (20-2)/2 + 2 = 11.

We define a *knowledge ratio* as the number of traits in the student's concept also in the expert's concept, divided by the count of traits in the expert's concept. Since the student thought of three of the five traits of median, the knowledge ratio is 3/5 = .6.

It's tempting to define an error ratio of (1 - the knowledge ratio) or .4. But that estimates only part of the student's error, the traits the student ignores that should be included. It does not capture the student's error in mistakenly adding traits to those that are correct. We address this below.

**Order**

One person A often knows what's involved in a situation or problem -- its elements or, as we say, its traits -- but is not clear about the requisite *order* or sequence of what's involved. Another person B knows both what traits are involved *and* the order of what's involved. In this situation, we say that B knows more than A. To assess this aspect of knowledge, we can deepen our idea of distance between concepts with a measure of order.

Consider the (set of) train stations between two locations, say Richmond, VA and Boston, MA. We can think of the stations with or without concern for the order as you pass them traveling on a train. If the person is not concerned with the order of the stations, thinking of

1) Richmond, Washington,D.C., *Wilmington, Philadelphia, Baltimore,* New York City, Boston

will do, but if the person is concerned with order, thinking of 1) (whether from left or right) will not do -- it shows that the person does not know the order completely.

The correct or expert's concept of the stations in order (from left to right representing travel south to north) is

2) Richmond, Washington, D.C., *Baltimore, Wilmington, Philadelphia,* New York City, Boston.

All stations are in both concepts, but 1) mistakenly thinks the order …Wilmington, Philadelphia, Baltimore, … is correct.

To create an appropriate measure of distance between the concepts (not the spatial distance between the stations!), we can conceive of transforming 1) into 2) in any of several ways. We might A) simply in *one* mental step move the trait of *Baltimore* to the left of (or before) *Wilmington.* Or we might B) in *two* mental steps first move *Baltimore* to the immediate left of *Philadelphia*, and then move *Baltimore* again to the immediate left of *Wilmington.* (In computer science, method B is called a "bubble sort" since each misplaced element is imagined, like steam bubbles in boiling water, to repeatedly "bubble" up to the place of its adjacent element until it reaches its proper place.) Or we might conceive of still other ways to effect the correct order. (In computing, implementation of this and other sorting methods may involve yet more steps.) These ways may count the distance increment for the order trait differently from the one of A and two of B. (Just as there are different ways to calculate the distance between two locations on the planet earth depending on the path traveled, there are different ways to calculate the distance between two people's knowledge concepts.)

I recommend method B because it recognizes the distance or amount a dislocated element is from its proper spot in the order. Method A, however, does not -- the conceptual distance increment for any dislocated element is always one, no matter how far it is from its proper spot. Whichever method we adopt, we should increment the expert's knowledge count by at least one for recognition of the trait or element of order.

**Ordering Actions or Events in Time**

Algorithms or formulas in math and computing often require that steps be followed in a certain order. In the typical algorithm to calculate a median and quartiles, the data values are first sorted and counted, and then the quartile or median values is counted off. Likewise, in the binary search algorithm in computing, the data values are sorted before the iterative procedure of searching for any desired value is begun.

In cooking recipes, expert chefs (as well as others) know that certain tasks (traits) must occur before others, if the dish is not to be a mockery of good food. In a cheesecake recipe,

1) The graham cracker crumbs and butter must be pressed into the baking pan *before* the filling of cream cheese, sugar, eggs, vanilla extract, etc. is poured into the pan, *not after*.
2) The pan must be filled with the ingredients *before* you put it into the oven, *not after*.
3) The topping must be put on the cake *after* it comes out of the oven, *not before*...etc…

The novice who puts the topping on the cake before baking it in the oven, or merely thinks of doing so, reveals a lack of knowledge of this one trait of switching the order of the two steps. The distance between the novice and the expert on this matter is at least one. If, in the style of hapless comedians such as Laurel and Hardy, the would-be chef also presses the crumbs and butter into the pan after the filling not before (so as to splatter the filling from the pan onto the floor and ceiling), the distance from the order 2 3 1 to 1 2 3 is two, since it requires conceptually, a la the "bubble" sort above, the two intervening reordering steps 2 1 3 and 1 2 3.

**Composition**

Any trait may be composed of others in turn. Molecules are composed of atoms, atoms of protons, neutrons, and electrons. Organs of tissues, tissues of cells, cells of nuclei, nuclei of atoms… The person who thinks of a set of traits as constituting another trait has a different concept than one who does not think of those constituent traits. The person whose concept agrees with the expert's concept – which may or may not share the constituent traits -- knows some things the other does not, and so has a higher knowledge count of the concept.

In a story or characters in a novel, play, movie, or song, details are stated, implied, or omitted. To provide details is to reveal the composition of a trait. In the classic American song, "Frankie and Johnny," "Frankie and Johnny were lovers…" but Johnny was not "true." Betrayal leads to revenge. If one knows more detail of the trait "revenge", the two individual's concepts differ, and, other traits being equal, one has a higher knowledge count. One person may think of the lyric -- Frankie "took out her big forty-four" revolver, and shot Johnny "through that hardwood door." Another may think Frankie got revenge by calling the cops or, as Jasmine did in Woody Allen's movie "Blue Jasmine," the FBI. The concept distance is 3. If the expert agrees with the first person, her knowledge count is higher by 2.

One who knows the ingredients comprising the cheese cake filling above -- namely, vanilla extract and lemon as well as cream cheese -- knows more and has a higher knowledge count by 2 than one who does not. Knowing the relative amounts of these increases the knowledge count and may increase the differences in traits between two person's concepts as well.

In many subjects, the steps in an argument are often omitted or implicitly stated. In analytic geometry, the proof for the formula for Euclidean distance between two points is based on the Pythagorean theorem that in a right triangle, the sum of the squares of the two sides equals the square of the hypotenuse. One who knows details of the Pythagorean proof knows more of the proof of the Euclidean distance formula.

In computing, one routine or procedure calls another to do something (perform a task), but the calling routine does not see the details of what the called routine does to complete the task. The sort algorithms mentioned above are usually implemented in such called routines. The calling routines are like unknowing investors who give their savings to a trusted financial advisor, saying "I don't care how you do it, just bring me a 7% return on my investment." But to determine the trait count of the called rule or procedure as more than one, we must have some ideas of the details of what the called routine does -- we must think of or know its composition to count the traits composing it.

The steps in a process or algorithm often follow rules or reasons that purport to justify them. The steps can be thought of as composed of the rules that justify them. Sometimes these rules are explicit as they are in elementary arithmetic or may be in propositional logic. Other times, the rules are often unstated or simply understood, as often in cooking,  or not able to be stated clearly at all, as in first speaking one's native language.  (Some researchers have focused on the importance of rules for test diagnosis. (Tatsuoka, K.K.et al.)) *The student who thinks of or states the relevant rule has a different concept, and, other things equal, a higher knowledge count, than one who does not.*

**Summary Statistics**

To promote estimates of knowledge and concept differences and distances, we define:

The *expert count* is the number of traits in the expert's concept.
The *student count* is the number of traits in the student's concept.
The *knowledge count* is the number of traits in the student's concept also in the expert's concept.

Figure 2 illustrates the ways in which two concepts --the student's and expert's -- can overlap and differ. The overlapping region in any Figure 2 diagram is the student's knowledge region.

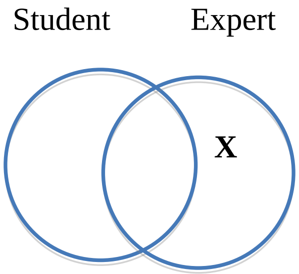
The *knowledge ratio* is the student's knowledge count divided by the expert count, or equivalently, the cardinality of the intersection of the two concept sets divided by the cardinality of the expert's concept set. This is a number between 0 and 1 inclusive.

The *ignorance count* is the number of traits in the expert's concept *not* in the student's concept. (The rightmost region in any Figure 2 diagram is the student's ignorance region.)

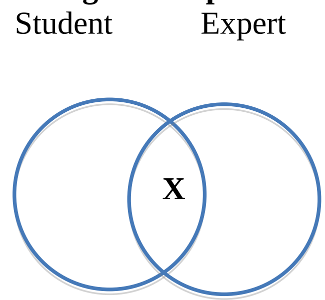
The *ignorance ratio* is one minus the knowledge ratio. Equivalently, it is the ignorance count divided by the expert count. This is also a number between 0 and 1 inclusive.

The *confusion (or fog) count* is the number of traits in the student's concept that are *not* in the expert's concept. (the leftmost region in any Figure 2 diagram.)

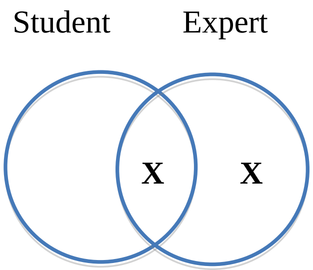
**Figure 2**
**Possible Relations Between Two Instances of Knowledge Concepts**



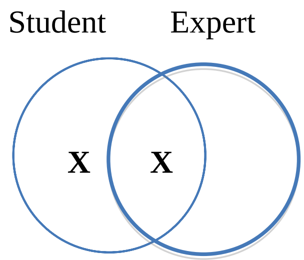

**1 Student doesn't know concept at all**



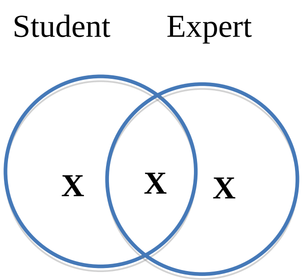

**2 Both know all**



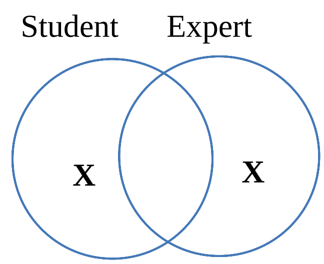

**3 Student knows some but missing some**



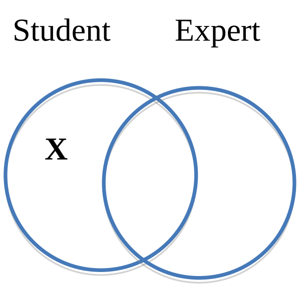

**4 Student knows all but wrongly adds some traits**



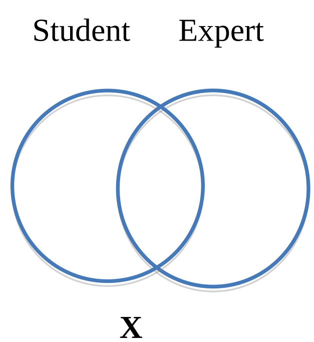

**5 Student has some not all and wrongly adds**




**6 Student has wrong traits, none right**




**7 Student has concepts Expert lacks**




**8 More things in Heaven and Earth, Horatio…**

Depending on our psychological assumptions about people in general, the specific students, and the particular knowledge concepts before us, we may find that the confusion count, in conjunction with the knowledge count and knowledge ratio, may be more informative in assessing student knowledge than the knowledge statistics alone. The human mind can focus only on a finite number of traits or ideas at once. If traits comprising the confusion count distract, influence, or, worse, overwhelm, the student's knowledge traits (the overlapping area), then the knowledge count and knowledge ratio seem compromised, since they do not indicate the confusing effect on the student's knowledge. (Figure 2 cases 4, 5,and 6.) To estimate this possible effect, we define

The *confusion or fog ratio* is the confusion count divided by the number of traits in the *student's* concept. This is a number between 0 and 1 inclusive.

Where the intersection is identical with both the student's and expert's concept set (the set differences are both null), the knowledge ratio is 1 and the confusion ratio is 0 (Figure 2 Case 2). This means that the student has exactly the same knowledge of the concept as the expert. Although all of these statistics apply to all 8 cases listed in Figure 2, some are more informative than others for specific cases. The knowledge and ignorance ratios, for instance, are informative for case 3. The confusion factor is informative for cases 4, 5, and 6.

In Figure 2 cases 4 and 5, the confusion ratio may or may not illuminate the significance of the knowledge ratio. If the confusion count is "small" relative to the student count (and so the confusion ratio is small), then the effect of these traits on the knowledge count and ratio is presumably small and we may choose to ignore it.

If, however, the confusion count is "large" relative to the student count (and so the confusion ratio is large), then the effect of these confusing traits on the knowledge count and knowledge ratio is possibly large. In that case, to portray the knowledge in a more accurate way, we should probably speak of the confusion ratio in the same breath that we speak of the knowledge ratio, or try to combine the two ratios in some informative way into a somewhat *"diminished" knowledge ratio*.

To combine them into one measure enables apparently simple comparisons between any two individuals' knowledge concepts, but it risks concealing differences in confusion counts between the two knowledge concepts.

How, then, could we combine the knowledge ratio with the confusion ratio? It may be tempting either to reduce the knowledge count by the confusion count and recalculate the knowledge ratio, or simply subtract the confusion ratio from the knowledge ratio. But there are several objections to either way of computing this "diminished" knowledge ratio.

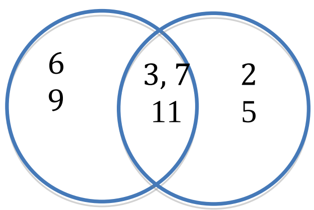
First, combining the two suggests that each confusion trait in some way *cancels or removes* a knowledge trait from the student's concept in his or her mind. Depending on the student, this may or may not be true. When it's not true, the combination statistic is misleading.

Second, it will result in negative knowledge ratios when the confusion count is greater than the knowledge count. This negative signals loudly that it is a diminished knowledge ratio -- prodding us to wonder what the original knowledge and confusion ratios actually are.

Third, unless the knowledge ratio is negative, it does not even begin to distinguish between the original knowledge ratio before it is "diminished" and after.  Two examples illustrate these difficulties.

Suppose, in pondering the prime numbers less than 12, one student thinks of 3, 6, 7, 9, 11. (Figure 2 case 5.) This student knows three of the five primes <=12 (namely, 3,7,11) for a knowledge count of 3 and a knowledge ratio of 0.6. (We're ignoring whether the students know how to decompose "prime number" into its definition.) The student's confusion count is 2 (thinking 6 and 9 were prime), and confusion ratio is 0.4.  The *"diminished" knowledge ratio* subtracting the confusion count is *(3-2) = 1 / 5 = 0.2.*

**Student 1 Expert** **Primes <= 12** **Student 2 Expert**



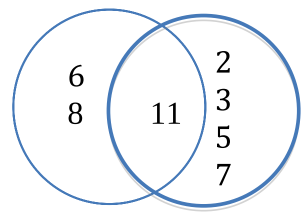




**Fig 3 Two Student Primes Concepts**

Another student (also Figure 2 case 5) thinks of the numbers 6, 8, 11 under the concept "prime <=12."  He thinks of only one prime 11 of the five (2,3,5,7, and 11) in the expert's concept, so his knowledge count is 1 and *(original) knowledge ratio is 1/5 = 0.2.* (The student's confusion count and ratio are 2 and .67). Without explicitly saying which it is, it is unclear whether this 0.2 is an original knowledge ratio or a "diminished" ratio as with the previous student. Without this distinction before us, we will likely not appreciate the greater knowledge and higher original knowledge ratio of the first student.

The problem is similar in other fields like sports, where we have one or more positive measures of achievement, and one or more negative measures of missteps, but no way to combine them. In baseball, we have batting averages and number of strikeouts. In football, we have number of pass completions and the number of interceptions. In both sports, there is no reasonable way to combine the opposing pair. So with knowledge counts and confusion counts and ratios/percentages, we should follow a similar policy and not combine the knowledge and confusion ratios. Rather, we should present both together.

## Dynamically Changing Knowledge Ratios

Our judgments of the knowledge of others (whether students or experts) can change as the current expert standard itself changes.  With an overlapping three-circled diagram, instead of two, we can represent the situation for a time interval where the expert's knowledge changes and the student's knowledge for the moment stays the same. The third circle represents the expert's knowledge at the later time t2. The situation becomes more complicated because each area of the two-circled diagram is now itself divided into two -- one comparing with expert's earlier concept, the other the later.

**Student Knowledge same,**
**Expert Knowledge increases**

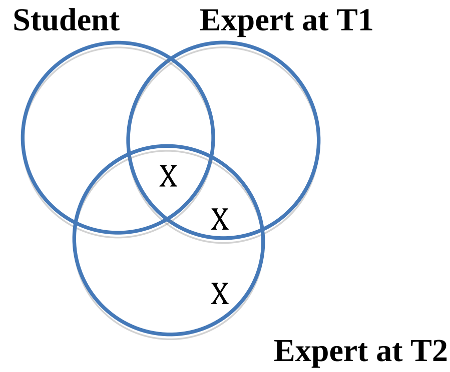



**Figure 4**

If the expert's knowledge increases overall, while the student's knowledge concept stays the same, the student's knowledge ratio will decrease, with the larger denominator. Conversely, if the student's knowledge increases, while the expert's concept stays the same, the student's knowledge ratio would increase.

## Conclusion

To varying degrees, we recognize concept traits, differences, and distances in ourselves and others, and roughly estimate their numbers. Through more attentive informal and formal observations of others' remarks, writings, and actions, we can deepen these abilities and better judge concept distances, knowledge ratios and other statistics. Pursued diligently, this approach will enable us to make more reasonable claims about one student or layman knowing more or less than another, and provide another mechanism for continual assessment of knowledge.

## References

Courant, R. and Robbins H. *What is Mathematics?* Oxford, 1946.
DeMorgan, Augustus. *Symbolic Logic*, Forgotten Books, 2012.
Frege, Gottlob. *The Foundations of Arithmetic*, tr. J L Austin, Basil Blackwell, 1968.
Katzen, Mollie. *Moosewood Cookbook*, Ten Speed Press, Berkeley, CA, 1992.
Mendelsohn, Elliot. *Introduction to Mathematical Logic*, Van Nostrand, *1964.*
Secolsky, Charles, and Levy, Stephen H., et. al. "Improving Validity By Assessing Students' Thought Processes" *National Council on Measurement in Education Conference*, 2013.
Tatsuoka, K.K, et al., "Cognitive Diagnostic Method: Rule Space, Q-Matrix Theory and Applications" in *Handbook on Measurement, Assessment, and Evaluation in Higher Education*, ed. Secolsky, Charles and Denison, D. Brian, Routledge, London, 2012.
Voelker, David, and Orton, Peter. *Quick Review Statistics*, Cliffs Notes, 1993.
Wittgenstein, Ludwig. *Philosophical Investigations*, London, 1952.
Young, Forrest W. *Multi-Dimensional Scaling*,
http://forrest.psych.unc.edu/teaching/p208a/mds/mds.html

## Acknowledgments

I would like to thank Charles Secolsky for insightful comments on the paper, and Beryl W. Levy for cultivating my appetite for the arts of cooking as well as their culinary effects.