13th International Congress on Mathematical Education
Hamburg, 24-31 July 2016

HIGH SCHOOL MATHEMATICS TEACHERS' UNDERSTANDING OF INDEPENDENT EVENTS

Adam Molnar

Oklahoma State University

Probability computations for independent and dependent events are included in many school curricula, but multiple meanings of the term independence make the problems challenging. As part of a study on teacher knowledge of probability in the US Common Core standards, 25 high school mathematics teachers were asked if two events in a two-dimensional table were independent. Only three of the teachers gave a correctly explained answer. None of the common errors – confusing independent with mutually exclusive, confusing independent with subsets, computational errors, and defining independence as lack of causative effect – were unforeseen. However, the confidence many study participants had in their incorrect answers was troubling. These results suggest teachers need support in order to effectively teach about independent events.

INDEPENDENT EVENTS IN SCHOOL CURRICULA

In probability and statistics, the concept of independence is fundamental. The definitions of independent and dependent events have had approximately the same meaning since De Moivre wrote in the 1700s. De Moivre wrote that two events are dependent if the "probability of either's happening is altered by the happening of the other" (De Moivre, 1756, p. 6); two events are independent if the occurrence of one event does not affect the probability of the other. For a modern definition, recently approved school academic standards in the US state of Oklahoma define dependent events as "events that influence each other. If one of the events occurs, it changes the probability of the other event." (Oklahoma Department of Education, 2016, p. A.4)

Probability computations for independent and dependent events are included in the upper grades of many school curricula: Grade 8 in Alberta, Canada (Alberta Education, 2014); Grade 7 – 10 (O-levels) in Singapore (Singapore Ministry of Education & University of Cambridge Local Examinations Syndicate, 2014); Grade 9 – 10 (Key Stage 4) in England (Department for Education, 2014); and Grade 11 – 12 (Level 8) in New Zealand (New Zealand Ministry of Education, 2007). In the United States, most state governments have adopted the Common Core State Standards in mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). In the Common Core standards, at some point in Grades 9 to 12, all students are expected to solve problems using the multiplicative definition that "two events A and B are independent if the probability of A and B occurring together is the product of their probabilities," as well as the conditional definition by interpreting "independence of A and B as saying that the conditional probability of A given B is the same as the probability of A, and the conditional probability of B given A is the same as the probability of B." (p. 82). Because high school students are expected to learn about independent and dependent events, their instructors need a solid understanding of the concept. Before describing the study undertaken to investigate teachers' knowledge, the next section of this paper details one major complication, the multiple meanings of the term *independent* in mathematics and statistics.

MULTIPLE MEANINGS OF INDEPENDENCE

Although De Moivre's definitions of independent and dependent events appear straightforward, the word *independent* and its related form *independence* have many other meanings. In everyday language, the first Oxford English dictionary definition is something "not depending on the authority of another, not in a position of subordination or subjection; not subject to external control or rule; self-governing, autonomous, free" ("Independent," 2016). Authority has a stronger causal implication than alter or change. In probability, one event does not have to directly influence the other to make the events dependent, but authority implies direct effect.

In mathematics, school-teachers in the United States frequently refer to *independent variables* when introducing algebraic functions. Common Core standards specify that Grade 6 students should "write an equation to express one quantity, thought of as the dependent variable, in terms of the other quantity, thought of as the independent variable" (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010, p. 44). An independent variable in algebra precisely specifies the value of the dependent variable, a strong level of causation similar to the dictionary sense, not the probability definition. Advanced mathematics includes more definitions of independence such as linearly independent vectors, but these uses rarely appear in pre-college mathematics.

Within statistics and probability, the word independence is used in at least three additional contexts besides discrete events. Independence is applied to continuous random variables, with a similar multiplication test in a different non-discrete context. In regression analysis, predictor variables are often called independent, similar to algebraic variables but with unknown error preventing exact specification. Data collection problems sometimes refer to *independent samples*. For instance, New Zealand students must recognize "the variability and independence of samples" (New Zealand Ministry of Education, 2009, p. 45). The description of independent samples is the same as independent events—values from one group provide no information about values in the other group—but independence is evaluated by knowledge about how the data was collected, not a formula. Guidance about that determination is frequently sparse. For instance, one common college introductory statistics textbook offers only the following definition: "Samples are independent samples when they are not related" (Bluman, 2014, p. 499). A mathematical statistics textbook (Wackerly, Mendenhall, & Scheaffer, 2008) properly notes that a completely randomized design leads to independent random samples (p. 653), but it does not provide a more general definition. Also, this statement appears in Chapter 12, after confidence intervals and hypothesis tests for independent random samples appeared in Chapters 8 and 10.

Students might see at least six definitions for independence: everyday non-mathematical, algebraic variables, probabilistic events, probabilistic random variables, regression model variables, and data collection samples. Collegiate textbooks sometimes utilize all six (Wackerly, Mendenhall, & Scheaffer, 2008). Secondary school students in the USA hear about everyday, algebraic, probability event, and sampling applications of independence. The level of lexical ambiguity is extremely high.

STUDY METHODS

A question about independent events was included in a doctoral dissertation on teacher knowledge and views about conditional probability in the Common Core standards (Molnar, 2015). In the study, 25 high school mathematics teachers from three US states (Georgia, Pennsylvania, and South Carolina) were asked nine task-based questions. At the time of the interviews, all three states included conditional probability topics in high school, but in different courses. Georgia included the topic in Analytic Geometry, primarily taken in Grade 9 or 10; Pennsylvania in Algebra II, primarily taken in Grade 10 or 11; South Carolina in courses for Grade 12 students.

During face-to-face interviews between May and July 2014, participants solved the questions, identified potential student misconceptions in each question, and suggested responses to misconceptions that would help students learn. The sample was acquired through personal and professional connections; it cannot be considered representative. In the group of 8 males and 17 females, 19 teachers held at least a masters' degree, primarily in education. All participants except one had taken at least one statistics course at university: 9 took one course, 6 took two, 6 took three, and 3 more than three. The sample contained 7 experienced probability instructors, 3 in the Advanced Placement (AP) program and 4 in non-AP courses, meaning that about three-quarters of the teachers had not taught a course on probability and statistics. Only 9 of the 25 participants had completed any professional workshops on probability and statistics teaching.

The question about independent events was taken from a 2010 AP Statistics examination (The College Board, 2010). The question presented the data shown in Table 1, about the primary source of news and level of educational achievement in a hypothetical random sample of 2,500 adults. The participants were asked "When selecting an adult at random from the sample of 2,500 adults, are the events 'is a college graduate' and 'obtains news primarily from internet' independent?"

Primary Source for News	Not High School Graduate	High School Graduate but Not College Graduate	College Graduate	Total
Newspapers	49	205	188	442
Local television	90	170	75	335
Cable television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1,437	693	2,500

Table 1: Data for independence question (The College Board, 2010).

The two events are not independent. This can be shown with either the multiplicative definition or the conditional definition. In the multiplicative approach, the solution involves testing if the product of the probability of being a college graduate (event C) and the probability of obtaining news from the Internet (event N) equals the joint probability of the two events. Symbolically, the test asks if $P(C) * P(N) = P(C \text{ and } N)$. Since $(693/2500) * (687/2500) \neq (245/2500)$ because $.076 \neq .098$, the events are dependent. Alternatively, dependence can be demonstrated with the conditional definition, since the unconditional probability of obtaining news from the Internet of $687/2500 = .275$ does not equal the conditional probability of $245/693 = .354$.

RESULTS

Of the 25 participants, only 3 gave a correct answer with a correct explanation. Table 2 lists the solution path given by each teacher, noting whether the attempt was correct or incorrect. Only 6 participants applied any mathematical formula; three-quarters of the teachers offered an incorrect non-numeric explanation.

Solution path	Correct	Incorrect
No, not mutually exclusive	0	10
Yes, there is no subset	0	6
Multiplication formula	2	2
Yes, lack of effect	0	3
Conditional formula	1	1

Table 2: Solution paths for independence question.

The most frequent misconception was claiming that the events were not independent because they were not mutually exclusive, P(N and C) > 0. As a teacher incorrectly stated, “I’m going to say no they are not independent groups, because I can be in both groups at the same time.” There were two types of incorrect statements in which participants concluded that the events were independent. In the answer labeled *subset*, teachers claimed independence because neither group was a subset of the other group, “because there are people who received news from the Internet who are not college graduates” as one teacher said. Symbolically, this claim is P(N and not C) > 0. In the other answer, teachers talked about the lack of *effect* of one variable on the other, but did not refer to subsets of full groups. These explanations referenced the everyday definition including causality. For example, one teacher said the events were independent because “it's the same as probability of having blond hair and the probability of being a girl, they’re two so totally different things. One doesn’t rely on the other. … Having blond hair has nothing to do with having or being a girl.”

Many of the teachers seemed to have confidence in their verbal explanations; fewer participants asked about the correct answer on this problem than on other problems. Explanations often were short and to the point. For example, one teacher said yes “because there are people who received news from the Internet who are not college graduates.” Another teacher explained a no answer with just three words, “because they cross.” When pressed for further detail, she said she explained the concept to her students through the song “Miss Independent” by Kelly Clarkson. She said, “I’d say what does she sing about, what does that mean, you know being on her own.”

Some teachers mixed correct and incorrect language. One teacher gave a correct example of independent events, rolling a die and picking a playing card from a standard deck of 52 cards. Unfortunately, that teacher also incorrectly claimed that the events “draw a king” and “draw a diamond” were not independent because with the king of diamonds, both events could occur together. Another teacher used the idea of effect, a word often correctly used to describe independence, but said the events were independent because “one does not affect the other,” incorrectly assigning a causative mandate to affect/effect.

Given subject matter results, it is likely not a coincidence that the most frequently mentioned student misconception was vocabulary, cited by eight teachers. When responding to vocabulary problems, teachers tended to suggest direct instruction, since without knowledge of the definition students cannot attempt to find an answer. Three teachers mentioned that students would have trouble with mathematical computation. Two teachers noted the confusion between independent and mutually exclusive events. Several other potential misconceptions were mentioned by one teacher; about one-third of the teachers made no misconception suggestions.

DISCUSSION

The high misconception rate on this question was very surprising, because the definition of independent events appears in many, many textbooks with sections on probability. In most books, such as Bluman (2014), the textbook definition is consistent with Common Core standards and De Moivre’s definition. During one interview in South Carolina, the teacher and interviewer looked up the correct definition in Chapter 10 of the teacher’s algebra book.

One possible reason for the high misconception rate was the lack of prominence given to probability in the curriculum. Across all three states, teachers mentioned that probability was a topic located at the end of the book, covered if time permitted. For example, a Georgia teacher said “we spent the first three months on the first unit, so we skimmed this unit. We didn’t even actually teach conditional … We had like two days to do it.” Although the South Carolina teacher had textbook support about independence, that year her class had only completed Chapter 8. One Pennsylvania teacher offered a farm analogy, saying “When we’re teaching to the Common Core standards for the state test, there might be one question on the whole bloomin’ test. This isn’t where you put your eggs, because you’ve got more eggs in other baskets.”

Another possible reason was the lexical ambiguity described earlier. During the conversation about misconceptions, one teacher illustrated the issue by discussing range, another term that has different meanings in algebra and descriptive statistics.

Teacher: Just getting this stuff mixed up. That’s not my input. What do you mean independent? I thought independent and dependent were input and output. Domain and range. And so that, that’s also very problematic for my kids when—

Interviewer: The wording with algebra is what you’re talking about.

Teacher: Yeah, algebra wording, yeah. So the algebra definition of independent as opposed to the statistical definition of independent, and kids expect, I mean, I’ve experienced that with range this year. When we talk about range with stats and creating box-and-whisker plots, they’re doing range. And so when I’m doing mixed review, and it could be anything, to go back to a range problem after just doing a stats problem [means] all right, What is range? And we have to go through the whole process again. So yeah, confusing vocabulary.

Lexical ambiguity is a compelling explanation for teacher misconceptions, because their answers were not wrong under other understandings of independence. The mutual exclusive explanation implies separation. The textbook examples for independent samples—completely randomized design and not related—can be considered separate. Informally, it is not uncommon to tell students to assume samples are independent because observations do not overlap in time or space. Asking people to discard this informal vocabulary of separation is, as the teacher said, confusing.

The subset explanation is related to algebraic and everyday definitions. In algebraic functions, the independent variable completely governs the value of the dependent variable. Having people outside the proper subset shows less than complete control. Under the mathematical operation of negation, not completely dependent equals at least partially independent. This reasoning does not lead to De Moivre's definition, but it is not incoherent. Similarly, citing a lack of causative effect as justification for independence, as three teachers did, is not illogical. That answer just uses non-mathematical understanding, not the statement for probability events.

This exploratory study was limited in scope to 25 volunteer participants, interviewed once with one question about independent events. Nevertheless, the high misconception rate and low level of doubt are worrying. It is unlikely that the general US high school mathematics teacher population has substantially higher levels of knowledge than a volunteer sample. If misconceptions are this prevalent in the teacher population (or even half as prevalent), accurate teaching of probabilistic event independence is not occurring in many classrooms. Because societies have decided to place probabilistic independence in school standards, teachers need adequate support in order to help their students learn about this fundamental topic.

The solution is not simple, because it involves more than practicing a formula. When teachers asked for the correct answer, they had no problems with computation. The multiplicative test, checking if P(C) * P(N) = P(C and N), is simpler than many other computations in school mathematics. Some students will struggle, as a few teachers mentioned, but techniques exist to help students learn to compute. Multiple meanings are the real struggle. The everyday "Miss Independent" definition with causation yields a wrong answer. The lack of clarity about independent samples, a topic frequently included in the same course as independent events, leads to confusion. Earlier use of independent and dependent variables in algebra further complicates the matter.

Renaming and careful defining can eliminate some lexical overlap. The regression meaning of independence is becoming less common, as descriptors *predictor* and *response* have replaced independent and dependent in an effort to not imply causation. When introducing functions in algebra, *input* and *output* might replace independent and dependent. Factors that make samples independent could be defined more carefully, with more than one example. Unfortunately, these strategies do not remove the disparity between the everyday causal definition and the probabilistic non-causal definition. Other words exist where people can navigate among multiple meanings. For example, the noun *box* can represent a three-dimensional container, a two-dimensional drawn rectangle, or dozens of other possibilities ("Box, n.2," 2016). To reach a similar state for independence will require great effort.

References

Alberta Education. (2014). *Mathematics: Kindergarten to Grade 9.* Edmonton, Canada: Government of Alberta. Retrieved from https://education.alberta.ca/media/8775377/k_to_9_math_pos.pdf

Bluman, A. G. (2014). *Elementary statistics: A step by step approach* (9th ed.). New York, NY: McGraw-Hill Education.

Box, n.2. (2016). In *Oxford English dictionary* (2nd ed.). Retrieved from http://www.oed.com/view/Entry/22297

The College Board. (2010). 2010 *AP Statistics Free-Response questions (Form B).* New York, NY: Author.

De Moivre, A. (1756). *The doctrine of chances* (3rd ed.). London, England: Millar.

Department for Education. (2014) *Mathematics programmes of study: key stage 4*. London, UK: Authors. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/331882/KS4_maths_PoS_FINAL_170714.pdf

Independent. (2016). In *Oxford English dictionary* (2nd ed.). Retrieved from http://www.oed.com/view/Entry/94325

Molnar, A. (2015). *High school mathematics teachers' knowledge and views of conditional probability*. Unpublished doctoral dissertation, University of Georgia.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards: Mathematics.* Washington, DC: Authors.

New Zealand Ministry of Education. (2007). *The New Zealand curriculum: Achievement objectives by learning area.* Wellington, New Zealand: Learning Media Limited.

New Zealand Ministry of Education. (2009). *The New Zealand curriculum: Mathematics standards for years 1-8.* Wellington, New Zealand: Learning Media Limited.

Oklahoma State Department of Education. (2016). *Oklahoma academic standards for mathematics 2016.* Oklahoma City, OK: Author. Retrieved from http://sde.ok.gov/sde/sites/ok.gov.sde/files/documents/files/OAS-Math-Final%20Version_2.pdf

Singapore Ministry of Education & University of Cambridge Local Examinations Syndicate. (2014). *Mathematics GCE ordinary level (Syllabus 4048).* Singapore: Singapore Examinations and Assessment Board. Retrieved from http://www.seab.gov.sg/content/syllabus/olevel/2016Syllabus/4048_2016.pdf

Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical statistics with applications* (7th ed.). Boston, MA: Cengage Learning.