



Finding a Happy Median: Another Balance Representation for Measures of Center

Lawrence M. Lesser

Amy E. Wagler

Prosper Abormegah

The University of Texas at El Paso

Journal of Statistics Education Volume 22, Number 3 (2014),
www.amstat.org/publications/jse/v22n3/lessert.pdf

Copyright © 2014 by Lawrence M. Lesser, Amy E. Wagler, and Prosper Abormegah, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Mean; Median; Measures of Center; Misconception; Representation.

Abstract

This paper explores the use of a lesser-known dynamic model for the median, a foundational topic that starts in the middle school curriculum and is associated with student misconceptions and knowledge gaps. This model appears to offer a rich vehicle to explore the median interactively in greater conceptual depth that includes some of its more subtle associated ideas. An exploratory study to assess performance of this model in a class for pre-service middle school teachers yielded evidence that students who completed the dataset sequence associated with the model gained further insight about the median, especially concerning how the mean and median are affected differently by outliers. Analyses of open ended questions as well as empirical results of multiple-choice questions are used to assess the overall learning outcomes gained by students. A one-minute video is offered to illustrate key points of the model.

1. Introduction

The mean and median are measures identified for learning as early as sixth grade in the Common Core ([National Governors' Association 2010](#)) and have representations included in the beginning level (Level A) of the pre-K-12 GAISE ([ASA 2007](#)). [NCTM \(2000\)](#) expects proficiency in determining and applying them by the end of eighth grade. Given how similar these words (and related words) look and sound (see Table 1 in [Lesser and Winsor 2009](#)) and how frequently they end up being introduced in the same day's lesson, it is conjectured that a more sustained

exploration of the measures that includes physical models can help give students a better conceptual and contextual basis for understanding and distinguishing the measures.

[Watier, Lamontagne, and Chartier \(2011\)](#) and its associated letter ([Lesser 2012](#)) catalog many conceptualizations of the mean: socialist (i.e., fair share or leveling value), fulcrum, algebraic form of least squares, geometric form of least squares, vector, and signal in a noisy process. However, there does not appear to be much identification and discussion of conceptualizations of the median beyond the middle value from a sorted list of discrete values or (in the continuous case) the value that divides a histogram or frequency distribution into two equal areas. This paper aims to present and discuss a lesser-known inexpensive physical model for the median that facilitates conceptual depth and yet is accessible for learners as young as middle school students. Preliminary implementation (described in Section 4 and [Appendix A](#)) suggests that some connections in this paper may be new and/or valuable even to college students, especially those who are pre-service teachers. To support classroom use, this new model for the median is illustrated in Section 4 with a sequence of classroom-ready, classroom-tested datasets in the spirit of existing task-based explorations of the mean ([Hudson 2012/2013](#); [O'Dell 2012](#)).

2. Balance Point Representations

2.1 Mean

ASA ([2007](#), pp. 41-43) expects learners at Level B (most often associated with middle school students) to recognize the mean as the dataset's "seesaw" balance point, a representation which [O'Dell \(2012\)](#) uses in an excellent sequence of conceptually-rich physical tasks. [Holmes \(2003\)](#) shows a model to facilitate transition to this balance representation from the Level A "fair share" notion. Other pedagogical discussions of the balance beam model include [Doane and Tracy \(2000\)](#), [Falk and Bar-Hillel \(1980\)](#), [Hardiman, Pollatsek, and Well \(1986\)](#), and [Hardiman, Well, and Pollatsek \(1984\)](#). The balance model has also appeared in textbooks (e.g., [Frankfort-Nachmias and Leon-Guerrero 2000](#)) and in a comprehensive collection of statistics mnemonics ([Lesser 2011a, 2011c](#)). Such representations may help students deepen their conceptual understanding ([Gfeller, Niess, and Lederman 1999](#); [Hardiman, et al. 1986](#)) and can be explored physically with measuring stick and fulcrum (e.g., [O'Dell 2012](#); [Penafiel and White 1989](#); [White and Berlin 1989](#)) or virtually with applets such as:

<http://www.shodor.org/interactivate/activities/PlopIt/>
http://onlinestatbook.com/2/summarizing_distributions/balance.html
<http://illuminations.nctm.org/Activity.aspx?id=3576>

2.2 Median

Some educators may assume the median does not call for as much exploration as the mean for one or more of these possible reasons: (1) the median is viewed by teachers or students as computationally simpler than the mean, (2) among measures of center, the mean "requires both additive and multiplicative reasoning" ([Hudson 2012/2013](#), p. 302), (3) the median (compared to the mean) has had less focus in terms of research about students' understanding and misconceptions (e.g., [Jacobbe 2012](#); [Pollatsek, Lima and Well 1981](#)), and (4) there has been more focus on summarizing interval data than ordinal data ([Mayén and Díaz 2010](#)). In any case, there is evidence that suggests more instructional time exploring the median could be helpful.

For example, [Groth and Bergner \(2006\)](#) found that only three of 45 pre-service elementary school teachers gave descriptions of the median that addressed both ordering as well as when the median is not a value in the dataset.

The need for a graphical model of the median is underscored by research ([Friel and Bright 1998](#)) indicating elementary school teachers' difficulty in finding the median from graphical representations of data. Also, [Cooper and Shore \(2008\)](#) discuss how students struggle to maintain understanding of measures of center (and variability) when data are represented graphically. [Leavy, Friel and Mamer \(2009, p. 346\)](#) describe and illustrate a model in which one folds in half a paper strip consisting of a row of equal-sized adjacent squares (with a data value in each one). This particular physical model, however, is limited in its ability to do "what if?" explorations of the effect of adding or deleting a particular observation.

Figure 1. Median Loop Balance Model for Dataset 2, showing student point of view.



[Lynch \(2009\)](#) developed an innovative dynamic physical model for the median involving a pulley wheel and a loop with equal-sized weights, using only ten dollars' worth of materials from a local hardware store. When we (the first two authors) developed our adaptation (see [Figure 1](#)), we used the following materials: a 4" diameter clothesline pulley wheel, a length of twine (about 6.5' long) to be tied into a loop, and a set of large office binder clips (whose clamping surface is 2" long and can accommodate a stack of paper 1" thick). The choice of binder clips (over, say, the hardware store spring-loaded clothespin-type mini-clamps used by Lynch; personal communication of Lynch with first author, August 11 and 21, 2011) was to make sure the weights were (1) more readily obtainable, and (2) heavy enough to overcome any friction that might unduly keep the loop at rest. Twine was used (instead of string) so that it was thick enough to add visible markings. The twine was then threaded through the pulley wheel and then tied into a loop. The unavoidable knot in the loop became an absolute landmark, which was chosen to be 17.5, a value close to (but not exactly in) the middle of the number line, and halfway between whole number values so that weights could be easily attached to the whole number on either side. Then, number line/loop markings were made 2" apart (which was conveniently the length of a binder clip) marked in black, with multiples of ten bracketed in red and odd multiples of five bracketed in blue (which can be remembered because "red" and "ten" have the same number of letters and so do "five" and "blue"). In this manner, a number like 22 could be distinguished readily from 27, for example. The bracket markings increased the size of the landmark, making multiples of 5 and 10 quick to find by size as well as by color. And knowing that the knot was at the value 17.5 allowed quick ascertaining of which direction corresponded to numbers increasing versus decreasing.

We conjecture that having a balance representation for the median as well as one for the mean can give students and instructors more tangible

appreciation for the concise observation of [Bock, Velleman and DeVeaux \(2007, p. 81\)](#) that the mean balances deviations and the median balances counts.

3. Pilot Assessment of Effectiveness of Intervention

3.1 Population

At a midsized urban southwestern US university with a predominantly Hispanic/Latino population, a statistics course for pre-service middle school teachers was offered, using the textbook by [Perkowski and Perkowski \(2007\)](#). There were 14 students enrolled in the course in the fall 2013 semester. On September 5, 2013, thirteen students attended and all 13 took the pretest. Ten of those 13 students attended the next class meeting (September 10) and all of those 10 participated in the activity and took the posttest, yielding 10 matched pairs.

3.2 Method

A pretest was administered that consisted of ten questions about the median. Because the first author was the instructor of this class, the pretest was administered and analyzed by other researchers so that he did not see any data until it was deidentified or aggregated. Also, it was made clear to students that they would not be graded on the pretest and that they had the choice to allow their responses to be part of the data set. There were 10 students present for the pretest and posttest and 100% of them gave consent. Then, at the next class meeting, after five minutes of “housekeeping,” students experienced a 45-minute demonstration (essentially, [Table 3](#) and [Appendix A](#) of this paper) and discussion of the balance models as part of a class meeting to explore “Measures of Center” (Section 3.1 of the course textbook). Thus, students had not discussed the mean and median in the course yet, though they very likely had some prior exposure to it before college because the topics are in the state’s K-12 school curriculum standards. The demonstration was then followed by 20 minutes to take the posttest (with the same questions and conditions as the pretest). During each step of the demonstration, the students were asked classroom voting questions (e.g., [Lesser 2011b](#)) and also had the opportunity to ask questions. At the beginning of the demonstration, students were given a handout of the sequence of datasets so that they could write notes on it and see which values were changed from step to step. The handout was basically the leftmost column of [Table 2](#) with the title “Price (in dollars) of DVDs of newly-released movies.”

3.3 Development of Assessment Instrument

There are ten questions on the pretest and posttest. Seven of these are multiple choice questions, two of which are followed by an “Explain” prompt to obtain an open-ended explanation. The remaining three questions are strictly open-ended questions with no multiple-choice component. This combination of open-ended and multiple-choice questions was designed to gain insight into student reasoning about the median before and after the activity. Potential questions were obtained by downloading items categorized as “Measures of Center” in the ARTIST database (<https://apps3.cehd.umn.edu/artist/>; [Garfield and Gal 1999](#)). The resulting item set was further narrowed by selecting what most closely aligned to the concepts in the dataset sequence. Wording was streamlined in items 1 and 7 in [Appendix B](#) and option C in item 6 was changed.

Item 10 was created from scratch by the first two authors to include a conceptual property of the median not found in the ARTIST collection. [Table 1](#) includes a brief description of each item as well as its aligned dataset(s), and the actual items appear in [Appendix B](#).

Table 1. Assessment items and their associated datasets.

Description of Item	corresponding dataset #
1: Provides an example of skewed count data and asks for reason why the mean would be inappropriate as a measure of center.	2
2: Asks for comparison of mean and median for symmetric data	1
3: Definition of median question	1-7
4: Asks for comparison of mean and median for skewed data	6
5: Find the median question with an odd number of observations (median is a value of the not evenly spaced dataset)	5
6: Find the median question with an even number of observations (median could be the mean of central two values of the not evenly spaced dataset)	7
7: Contextualized questions asking about a valid interpretation of a median	2
8: Asks for approximate value of median from a dotplot	1-7
9: Asks for comparison of mean and median for a dotplot	1-4
10: Asks for median from a given contingency table	1-7

4. Classroom Balance Model Exploration

Using this model, we present a dataset sequence that incorporates teacher prompts and student responses and forces engagement with conceptual properties of the median. This sequence below is actually a revised sequence that incorporates helpful feedback from: piloting among colleagues, receiving helpful feedback from anonymous *JSE* referees on the first version of this paper, and using it with students in an actual classroom (as noted in Section 3.2).

In light of the [ASA \(2007, p. 30\)](#) recommendation for Level A students, we start with a data set with an odd number of values. Consistent with recommendations (e.g., [ASA 2005](#); [Neumann, Hood, and Neumann 2013](#)) for using data with real-life context, we use data on the price (in dollars) of new-release movies on DVD. Prices are rounded to the nearest dollar because, as a practical matter, the physical balance model would not allow one to distinguish \$23.99 from \$24.

At the beginning, students received a handout of the sequence of datasets used so that they could readily follow the progress of the demonstration and take notes. The sequence appears in [Table 2](#) with the values of the mean and median added for reader convenience, though they were not on the student handout. The sequence was developed to explore the conceptual properties of the model in an efficient manner by minimizing the adjustments needed to progress from one dataset to the next.

Table 2. Sequence of datasets used in the exploration (the model for the mean was not used after the first four datasets).

Dataset	Sample mean, \bar{x}	Sample median, \tilde{x}
#1: {20, 22, 24, 26, 28}	24	24
#2: {10, 22, 24, 26, 28}	22	24
#3: {12, 24, 26, 28, 30}	24	26
#4: {12, 24, 24, 30, 30}	24	24
#5: {10, 11, 12, 24, 24, 30, 30}	not used	24
#6: {12, 24, 24, 30}	not used	24
#7: {12, 24, 30, 35}	not used	anything between 24 and 30

Some practical recommendations arose from rehearsing and experience that may not all occur to an instructor using the median balance model for the first time. The instructor needs to hold the pulley so that the side of the loop representing smaller values is on the left side from the students' view, which means on the right side from the instructor's view. The instructor should also be careful about language and tell the students that "high" or "low" will refer to actual numerical data values and that if the height of an object on the loop needs to be referenced, we will speak in terms of "elevation" or "closer/further to the wheel."

For juxtaposition, the fulcrum model for the mean was also used to model the first few datasets of the sequence. A 48-inch "enhanced yardstick" was used for this purpose because it was a convenient length and effectively already has a number line marked on it. Non-insulated alligator clips (about 2" long) were placed on the stick at the appropriate values of the dataset, but the stick was not placed upon the "fulcrum" (a tan rhombus from a set of "pattern blocks," as illustrated in [O'Dell 2012](#)) until it came up in the discussion. [Figure 2](#) shows how the clips can easily attach to the stick or (in the case of repeated values in a dataset) to each other.

Figure 2. Platform Balance Model for the Mean, illustrating Dataset #4.

[Table 3](#) outlines the datasets used and the steps to model the dataset on the loop-and-pulley model (MODELED DATA), the context of the data set (CONTEXT), and the concept being explored for the median activity (CONCEPT). There are important in-class assessment questions (that can be implemented with clickers or voting cards, as in [Lesser 2011b](#)) and prompts included in [Appendix A](#) that correspond to each of the modeled datasets in [Table 3](#). With the latter, we supply instructor prompts (using italics for actual words an instructor might say), example classroom voting questions that can be used (and the results the first author obtained when using

them during this study, with the correct answer in boldface), and targeted observations or insights the students are encouraged to have. Note that answer choices to the questions the instructor posed often have rhyme (e.g., “A: it will stay”) or alliteration (e.g., “B: it will balance”) to make it easy for students to remember the choices long enough to give an answer without the questions having to be visually written or displayed. While the task sequence of [Table 3](#) and the assessment in [Appendix A](#) are presented separately, the assessment is no less important in the activity sequence.

For instructors who do not want to take the class time to go through all seven datasets or who do not want to take the personal time to build the pulley-and-loop physical model in the first place, we offer the resource of a one-minute video that shows some key properties in real time. The video can be found on the JSE site at http://www.amstat.org/publications/jse/v22n3/pulley_loop_physical_model_of_median.html or on YouTube at <http://www.youtube.com/watch?v=-dDbWTPL0vE>.

Table 3. Activity sequence for the median exploration.

Modeled Data	Context	Concept
Dataset 1: {20, 22, 24, 26, 28} While holding the string tight, place the values (2" binder clips) at: 20, 22, 24, 26, and 28. There is no pressure on the instructor to get the placement exact because with the median, the key is position, not precision. ^a	The data represents prices (rounded to the nearest dollar) of a sample of new-release movies on DVD. The pricing also reflects how most new releases are priced at distinct price points, but often reasonably similarly.	Odd number of values, equally spaced
Dataset 2: {10, 22, 24, 26, 28} While holding the string tight, move the clip at 20 to 10.	One DVD is given a special sale price (\$10) to attract people to the website/store, a use of the “loss leader” strategy to spur overall sales.	This dataset has an outlier (and it is probably better that it is a “low outlier” because students may be more used to seeing examples where the outlier is high).
Dataset 3: {12, 24, 26, 28, 30} While holding the string tight, add 2 to all values. The instructor has a choice of how to do this. The longer (but more explicitly transparent) way is to move each clip to a position with a value two units higher. Alternatively, the end result can be obtained in two moves instead of five by moving 10 to 12 and then 22 to 30.	An across-the-board price hike adds \$2 to all DVDs.	A dataset is formed by applying a location shift (i.e., a translation) to the previous dataset.
Dataset 4: {12, 24, 24, 30, 30} While holding the string tight, move the 26 down 2 units to 24, and move the 28 up 2 units to 30.	The price structure has been streamlined so that the five DVDs have only three different price values instead of five different values.	Dataset whose ordered values include repetition. In this particular case, the value of the median is one of the repeated values and the value of the median is not affected by whether or not frequencies of the observations are considered. ^b
Dataset 5: {10, 11, 12, 24, 24, 30, 30} While holding the string tight, add clips to 10 and 11.	Two more DVDs come out and are put on sale at a very special low price (\$10 and \$11) as a further use of the “loss leader” strategy to spur overall sales.	Another dataset whose ordered values include repetition, but this time the value of the median is affected by whether or not the frequencies of observations are taken into account. ^c
Dataset 6: {12, 24, 24, 30} While holding the string tight, remove the 10 and 11; also, take off one of the 30s, so we now have an even number of values.	The \$10 and \$11 DVDs and one of the \$30 DVDs sell out or is out of print, so they are gone from the company’s inventory, leaving just four DVDs in the dataset.	There is an even number of ordered values and the numbers in the two middle positions are equal.
Dataset 7: {12, 24, 30, 35} While holding the string tight, take one of the clips at 24 and move it to 35.	Demand increases for one of the \$24 DVDs, and so its price is raised to \$35.	An even number of ordered values again, but this time the middle two values are distinct.

^aIt is best before the clamps are added that the string is not positioned with the median already at bottom.

Ensuring this shows students that the pulley is truly free to move and is not unduly hampered by friction.

^bThe median of the set of observations equals the median of the multiset.

^cThis time, the median of the set of observations does not equal the median of the multiset.

4.1 Follow-up to Dataset #2: Modeling the Effect of an Outlier

In the likely event that at least one student desires or requests scientific insight or intuition into the result of the physical model, the instructor can be prepared to relay the observation of E. Loewenstein (personal communication with first author, August 11, 2011):

“the position of the weights along the loop doesn’t affect the balance per se; only the total amount of weight on each side matters, since all of the weights on the same side act essentially with the same moment. Equal amounts of weight on each side of the loop exert equal forces on the two sides, because all of the force is along the rope since the rope is vertical. So we only require the same number of [equal] weights on the two sides in order to balance. The only thing that matters is which side of the ‘fulcrum’ the weight is on.”

Students can model this last sentence by taking another two clamps and putting one anywhere on the left side and the other anywhere on the right side, and seeing that the value of the median is preserved. In Section 6.3, we discuss why the weight of the loop for the median balance model does not matter, but the weight of the beam for the mean balance model does.

An instructor can follow up Dataset #2 by discussing an additional real-world example of the importance of this property such as finding a measure of center of household incomes or house prices in a city, distributions that are generally right-skewed. Also, the instructor can have students calculate the mean of this new data set to show that the values of the mean and median can be different.

Showing that the median is not affected by the exact position of the outlier – whether it is at 15, 20, 25, etc. – is an opportunity to highlight another difference between median and mean: it is possible to find the median, but not the mean, of a set of numbers even if the exact value of the maximum is unknown. For example, suppose we have salary data that Person A earns \$40,000, Person B earns \$60,000, and Person C earns something over \$1,000,000. We cannot calculate the mean, but we can still state the median. The unknown value or values of a dataset need not be outliers. For example, from College Board (2013), we have this real-life data on college-bound high school seniors who took the SAT test:

Table 4. Years of mathematics studied by college-bound high school seniors.

# of test-takers	19,761	14,080	29,292	152,717	899,396	169,610
# of years of study in mathematics	½ or less	1	2	3	4	over 4

Because of the open-ended categories at both ends, 15% of the exact values are unknown, and the mean cannot be computed exactly. We can, however, identify the median because a dataset with 1,284,856 values has a median between the 642,428th and 642,429th positions. Counting from either end of [Table 4](#), it is not hard to see from the table that the middle of this dataset will fall among the 899,396 students who answered “4,” and so the median equals 4. Students can probably suggest other variables (e.g., age, income, etc.) often on surveys where the top category choice provided would have no ceiling.

4.2 Follow-up to Dataset #7: Implications of Non-uniqueness

When the string is released, ask students to “Find the low point of the loop and decide whether it’s a unique value.” Students are surprised to discover that the loop actually balances with the low point being *any* point between the two middle values 24 and 30. We note that, in Dataset #7 of [Appendix A](#), nine students chose choice B, which is consistent from the textbooks they have likely seen before, including the textbook (p. 50) of this current class. This nonuniqueness is intriguing to some students and disconcerting to others, even if they remember that a mode may not be unique (e.g., some distributions are bimodal). While graphing calculators and many textbooks (including the one the students in our study used: [Perkowski and Perkowski 2007](#), p. 50) usually adopt the convention to let the median be uniquely defined as the mean of the middle two values (two values playfully called comedians by [Stigler 1977](#)) in the case of the dataset having an even number of values, they do not generally explicitly acknowledge that it is a convention rather than an inherent property of the median. A rare exception is [Han, Kamber, and Pei \(2012\)](#) who say (referring to a particular dataset):

“There is an even number of observations (i.e., 12); therefore, the median is not unique. It can be any value within the two middlemost values of 52 and 56 (that is, within the sixth and seventh values in the list). By convention, we assign the average of the two middlemost values as the median.” (p. 46)

Indeed, this convention does not at all obviously follow from the looser description in books (or p. 29 of [ASA 2007](#)) that “The same number of data points (approximately half) lie to the left of the median and to the right of the median.” Such a description is also consistent with how some books describe the median M of a distribution as any number (i.e., a number that is not necessarily unique) M satisfying both of these conditions:

$$P(X \leq M) \geq .5 \text{ and } P(X \geq M) \geq .5. \quad (\text{Lindgren 1976, p. 57}).$$

The loop demo gives a physical basis for the latter (“*a* median”) conceptualization and against being bound by the former one. Many statistics books (including the one used by the subjects of this study) give instructions to “find the median” rather than “find a median.” Perhaps textbook authors make this choice so that their exercises will have unique answers (craved by most students) and so that they can give collective directions (as in [Perkowski and Perkowski 2007](#)) such as “For each of the following data sets, find the mean, median, mode, range, sample standard deviation, and sample variance.” (p. 81).

There are some reasons for a thoughtful examination of the convention. The Common Core ([National Governors Association 2010](#)) lists several examples of mathematical conventions that K-12 students must navigate, including: use of parentheses, order of operations, and x and y coordinate axes. Probability has conventions such as defining $0!$ to equal 1. Statistics has conventions, such as the typical use of Greek characters for population parameters and Latin characters (or Greek characters wearing a hat) for sample statistics. There is a convention to indicate with a break or squiggle when a vertical axis does not start at zero.

Another reason to explore the commonly (but not unanimously) used convention about the median is that a common theme in statistics curricula is making sure students are not misled by statistical summaries. Suppose there is a dataset with an even number of values such that there is

a large distance between the two middle values (this might happen in a small dataset such as Dataset #7 or even in a large dataset if the data is strongly bimodal). An advertiser or activist might report “the median” as one of the two middle values instead of as the mean of those two values. A concrete example of how different those choices might be is a dataset of ages of pedestrians in fatal accidents ([Roe, Shin, Ukkusuri, Blatt, Majka, et al. 2010](#)), which happen mostly to the very young and the very old, and such a dataset might have no values between 20 and 60 – a large window of possible medians! Most of the time, however, students will not encounter large real-life datasets with bimodality extreme (or U-shaped) enough where the difference between one of the middle two values and the mean of the middle two values will be of practical concern (indeed, if it did, the mean may be a more appropriate and robust measure of center than the median for this dataset in the first place).

The Common Core ([National Governors Association 2010](#)) requires sixth-grade students to encounter “mean absolute deviation” (MAD) as a measure of variability, where each deviation refers to the “distance” between a data value and the mean. It turns out that this mean absolute deviation would attain its least possible value if each distance were instead calculated from a median of the dataset, a fact which can be proven by induction, integral calculus, or even without calculus (e.g., [Schwertman, Gilks, and Cameron 1990](#)). This property of a median, together with two applications of Jensen’s inequality, can be used to prove the elegant but lesser-known result that the mean and median never differ by more than one standard deviation ([Mallows 1991](#)).

Even if students have not yet completed Algebra I, the fact that they have been working with “absolute deviations” since sixth grade should prepare them to be introduced to absolute value notation and appreciate the visual example of graphing the function involving the sum of the (absolute) distances between each value of the dataset $\{1, 2, 6\}$ and an arbitrary value x : $|x - 1| + |x - 2| + |x - 6|$. Let us relate this to the real-world physical context of anticipating which elevator from a bank of elevators will be the next to arrive (e.g., [Hanley, Joseph, Platt, Chung, and Bélisle 2001](#); also, see Hanley’s applet at <http://www.medicine.mcgill.ca/epidemiology/hanley/elevator.html>). Imagine a bank of six equally-spaced elevators numbered 1 through 6. Now suppose elevators numbered 3, 4 and 5 are out of service, and you are trying to figure out where (x) to stand to minimize the distance you walk before walking into the next elevator that arrives. (Assume all elevators are equally likely to be the next one to arrive and that when it does, you will walk in one dimension along the “elevator number line” until you are in front of the opened doors and then step straight in.) The expected value of the distance you will walk is $\frac{1}{3}(|x - 1| + |x - 2| + |x - 6|)$, which is minimized when $T(x) = |x - 1| + |x - 2| + |x - 6|$ is minimized. We can safely assume x is between 1 and 6, so let’s consider two cases:

(i) x is between 1 and 2, in which case $T(x) = 7 - x$, or (ii) x is between 2 and 6, in which case $T(x) = x + 3$. Notice that in both cases, $T(x)$ achieves its minimum value of 5 when $x = 2$, which is the median of $\{1, 2, 6\}$. Note that had the mean (3) been chosen, $T(x)$ would be $6 > 5$. [Hanley et al. \(2001\)](#) makes the insightful point that if one moves away from the median point, “one moves away from more elevators than one moves towards, thereby increasing the sum of the distances to the elevators.” (p. 1)

Now suppose that elevator #4 is repaired so that the elevators numbered 1, 2, 4 and 6 are now in service. So we want to minimize $U(x) = |x - 1| + |x - 2| + |x - 4| + |x - 6|$. If we consider three cases, we see that (i) when x is between 1 and 2, $U(x) = 11 - 2x$, which is minimized at $x = 2$; (ii) when x is between 2 and 4, $U(x) = 7$; (iii) when x is between 4 and 6, $U(x) = 2x - 1$, which is minimized at $x = 4$. Thus, any value of x between 2 and 4 yields $U(x) = 7$, the minimum value of $U(x)$. We conclude that standing anywhere between 2 and 4 will result in the same expected walking distance for the elevator.

Now have students enter the expression $U(x)$ into the $Y=$ menu of a TI-73 or TI-84 graphing calculator, accessing the absolute value function ABS from the top of the NUM menu after hitting the MATH key. Teachers should readily give students appropriate window settings (e.g., $X_{\min} = -5$, $X_{\max} = 10$, $X_{\text{sc}} = 1$, $Y_{\min} = 5$, $Y_{\max} = 20$, $Y_{\text{sc}} = 1$) so that the students can see that the graph is piecewise linear with a flat piece at the bottom. That horizontal segment shows the graph achieves its minimum value for any value x between 2 and 4, inclusive. This is also suggested by using the calculator to explore for that expression a table of values such as the simplified excerpt in [Table 5](#):

Table 5. Sum of absolute distances of x from each value of the dataset.

x	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5
$ x-1 + x-2 + x-4 + x-6 $	9	8	7	7	7	7	7	8	9	10	11	13

As an aside, this potential lack of a unique value for the median is one reason why the conventional line of fit (as used in the graphing calculator, for example) is not defined to minimize the mean absolute deviations of the data points from the line since such a line would not be unique ([Lesser 1999](#)).

To distinguish mean and median with respect to data type, the median loop balance model could be used to find the median of purely ordinal data, such as the problem presented in [Mayén and Díaz \(2010\)](#) with two datasets each consisting of letter grades (A, B, C, D) and the student charged with deciding which group obtained better scores and which central tendency measure would be useful. For each dataset, an unmarked loop can be used and clamps of four different colors can be attached to the loop in various places (as long as all the As are consecutive, followed by all of the Bs, then all of the Cs, then all of the Ds). From such an ordinal data set, where we do not know the exact values of the grades (e.g., 95 and 99 would both be lumped in the “A” category), we cannot compute a mean or represent the mean using a balance model. Or, to adapt the example to our DVD context, we could say that a pricing structure has been developed where the prices of five DVDs are represented by A, B, C, D and E, but all we have been told is that $A > B > C > D > E$. We know that the median of the movie prices is C, but we have no idea what the mean price is. Or perhaps A, B, C, D, and E could represent the quality (as rated by viewers) of the respective five DVDs.

5. Results

Ten items were utilized to assess student understanding about the median before and after the dataset sequence. The results of the seven multiple-choice items are summarized in [Table 6](#).

Recall that two of the open-ended questions are accompanied by a multiple-choice question. First an overall test of the proportions correct was conducted for all seven multiple-choice questions (thus, number of responses for pre and post groups was 70). There was a moderate effect on scores after the activity ($H_a: \mu_{\text{pre}} - \mu_{\text{post}} < 0$; $Z = -1.71$; $p = 0.0438$). A permutation test was also performed with resulting permutation *p-value* = 0.034. Thus, there is reason to investigate the questions individually after this overall assessment.

Due to the small sample size ($n=10$ respondents), permutation tests were utilized to assess whether the proportion correct increased for each individual questions. In [Table 6](#), the resulting permutation test pointwise and multiplicity adjusted *p*-values are provided. A Sidak correction was employed for the multiplicity adjustments. When assessed overall, it is apparent that the students gained knowledge about the relationship between the mean and median for skewed data. This demonstrates a practical as well as statistically significant effect. Other questions that demonstrate pointwise significance include the question assessing whether the student knows the median definition and asking for a valid interpretation of the median in a specific context. Given that 50% of the students could already pick out the correct definition of the median before the activity, the gain of two more students answering this correctly is arguably not practically significant. Three more students could pick out a correct interpretation of the median after the activity, perhaps due to the focus of the activity on the median being a central value.

Also included in [Table 6](#) is a column providing the proportion of students who answered aligned in-class polling questions correctly. All polling questions had some alignment with items 3 and 10 and are therefore left empty. It is interesting that some polling questions had high levels of correct responses in class but students consistently did not respond correct to an analogous item in the pre-post test (e.g., item 2). Otherwise, for item 4, there is strong evidence that students had a better concept of how the mean and median differ for skewed data using the pre-post test data. The in-class polling question backs this up, with 90% of students answering these questions correctly in class. With regard to items 5 and 6, there is little if any conceptual gain for these items (asking students to find a median for odd and even numbered sets of observations). The polling questions demonstrate that when students had the aid of the physical model, they understood how to find the median for an odd number of observations (80% correct), but did not recognize that the median need not be unique (the bisection between the middle values) for an even number of observations (10% correct). A full 100% of students were able to answer the polling questions of Dataset #2 correctly, but made only minimal gains when answering a question about valid contextualized interpretation of a median.

Table 6. Assessment results for multiple-choice items.

Item from Table 1	Proportion correct		Permutation test <i>p</i> -value	Adjusted <i>p</i> -value	Proportion correct for in-class polls (dataset where voting occurred)
	Pre-test	Post-test			
2	0.1	0.1	0.234	1.000	0.900 (#1)
3	0.5	0.7	0.077	0.539	0.678 (#1-7)
4	0.2	0.9	0.000	0.000	0.900 (#6)
5	0.6	0.7	0.290	1.000	0.800 (#5)
6	0.7	0.7	0.323	1.000	0.100 (#7)
7	0.3	0.6	0.031	0.217	1.000 (#2)
10	0.3	0.5	0.074	0.518	0.678 (#1-7)

Questions 2 and 4 on the pre-post test had multiple-choice responses followed by a prompt to give an (open-ended) explanation of the response chosen. The three remaining questions (#1, 8, and 9) are open-ended questions only. These were assessed by three statistics graduate students independently using a simple scoring rubric. The graduate students were accustomed to grading undergraduate student work in introductory statistics courses as grading is a major source of graduate student funding in the department. The rubric had four levels: (1) virtually no relevance, (2) some relevance, (3) substantively relevant, (4) completely relevant. Each student independently scored the student responses, and the inter-rater reliability (intraclass correlation coefficient (ICC) for multiple ordinal ratings; [Shrout and Fleiss 1979](#)) was computed for each question and at each time point. After the scoring by the three graduate students, ICCs were calculated and are given in [Table 7](#) in parentheses. Most of these indicate a reasonable level of correspondence between the scorers. [Table 7](#) also summarizes the mean score ratings for each question and time point, summarizing the nature of each question for space reasons (since many questions had accompanying figures, etc.). Note that question 4, assessing a comparison of the mean and median for skewed data, exhibits strong levels of inter-rater reliability and also shows a strong mean increase in the rater's mean scores. This result is consistent with the multiple-choice questions, where the proportion correct increases significantly for questions about the relationship between the mean and median. Also, this is consistent with the results of the in-class polling questions where all students were able to locate the mean and median accurately for skewed data after viewing the model of the median (see Dataset #2 of Appendix A).

Table 7. Mean ratings of student responses to open-ended questions.

Description of Question	Pre-test mean Likert rating (intraclass correlation coefficient)	Post-test mean Likert rating (intraclass correlation coefficient)
1: Provides an example of skewed count data and asks for reason why the mean would be inappropriate as a measure of center.	2.13 (0.101)	2.60 (0.194)
2: Asks for comparison of mean and median for symmetric data, then requests explanation	1.70 (0.291)	1.73 (0.253)
4: Asks for comparison of mean and median for skewed data, then requests explanation	1.47 (0.594)	2.37 (0.408)
8: Asks for approximate value of median from a dotplot	1.87 (0.147)	1.63 (0.607)
9: Asks for comparison of mean and median for a dotplot	1.60 (0.290)	1.93 (0.187)

6. Discussion

6.1 (De)limitations

We chose to use the population of pre-service middle school teachers for this exploratory study not just because they were convenient to access, but also because they provided an “upper bound” benchmark in that if they, as college students, struggled with the concepts, it would surely be all the more of a struggle for middle school students. The gap between these two groups of students may well be less than expected because of the university’s low selectivity index in admissions (about 99% of applicants are admitted) and higher than average age of student body (which means it would have been more years on average since students had seen some of the related material in precollege coursework). We assessed students from only one class because that class was the only section of the course that was offered in the 2013-14 school year. In fact, it was the first section of this particular course ever offered at the university in an attempt to meet identified probability and statistics content gaps in the preparation of pre-service middle school teachers, and the sample size (i.e., enrollment) for this debut class may have been lower than expected because word may not have gotten out to all potential students and advisors about the availability, nature, and importance of the course, which was offered using the course number of the “topics” placeholder course rather than being assigned a unique name and number. While this should be less of an issue in the future, it is a big reason our study can be considered only a pilot study and its findings not definitive.

6.2 Interpretations of Findings from the Dataset Sequence

The general insight students gained by performing the dataset sequence was how the mean and median are affected differently by outliers. The pre-post test data, in-class polling results, as well as the open-ended assessment questions back this up (see [Tables 4](#) and [5](#)). We note that this is

one of the most important learning objectives having to do with the median and is an essential insight for knowing when to report a median rather than a mean as a measure of center. Otherwise, using the physical model of the median, students were able to recognize that the median need not always be unique (see [Table 4](#)). However, it is unclear whether this concept transferred well to assessment questions where the physical model of the median was not there as a reference.

We recognize that one property was not reflected in the pre-post test: the potential nonuniqueness of the median. An example of an item we have developed for possible future research to assess this characteristic is:

Given the dataset {2, 18, 22, 26}, which of the numbers below could be reported as the dataset's median?

- A) 17
- B) 19
- C) 24
- D) More than one of the above
- E) None of the above

Students who believe the answer must be 20 will choose option E while those who recognize the median need not be unique for this setting would choose option B. Students who are confusing the mean and median would choose A and those mistaking the range for the median would choose C. Those recognizing the median need not be unique but not recognizing that the values must be between 18 and 22 may mistakenly choose D.

6.3 Additional Reflections

It is hoped that this model may help teachers feel more prepared for some of the pedagogical subtleties that can arise when teaching this apparently simple measure of center, as illustrated in the dramatic sketch by [Lesser \(2013\)](#). The importance of understanding how the mean and median differ is underscored by NAEP data ([Groth and Bergner 2006](#), p. 46): “On a test item containing a data set structured so that the median would be a better indicator of center than the mean, only 4% of Grade 12 students correctly chose the median to summarize the data set and satisfactorily explained why it was the appropriate measure of center to use.” That specific test low-performing item (with the context of attendance figures over five days for two movie theaters) can be viewed in item 3 of [Zawojewski and Shaughnessy \(2000b\)](#).

Additional real-world datasets can be explored (e.g., Appendix 1 of [Garfield 1993](#)) to give students more practice in choosing the better measure of center. Sometimes it is a matter of identifying whether a particular real-world dataset is unlikely to have much skewness or outliers (e.g., interest rates on 90-day certificates of deposit in major cities) or is likely to have them (e.g., incomes or housing prices). Other times, it is a matter of considering what question is of interest. For example, suppose you have a sports team and you know the salary of each player. If you want a sense of how much the typical player makes, you would want the median in case there is a superstar with an “outlier” salary. If you have a business interest in the team that depends on total payroll costs, then the mean is more relevant.

Students who experience placing clamps on a number line loop gain a concrete basis to visualize the data as *ordered* when finding the median in the future from a dataset on paper. This would be no small thing because students learning about the median often focus only on the number in the middle position and forget the prerequisite of seriation. Indeed, two-thirds of U.S. twelfth-graders could not find the median from unordered data on the NAEP ([Zawojewski and Shaughnessy 2000a](#)).

The balance models for the mean and the median share some pedagogical similarities, such as having connections to concepts from science such as force, moments, equilibrium, etc. Another similarity is that both balance models support “prediction” questions of what will happen when a particular value is changed or added to the data set. Yet another similarity in pedagogical properties is how both balance models can illustrate certain properties of the mean or median (e.g., [Groth and Bergner 2006](#); [Strauss and Bichler 1988](#)). For example, the fulcrum of the balance beam and the lowest part of the loop need not correspond to where a weight is located, which reinforces the idea that the mean or median can be but need not be one of the values in the dataset. Another property indicated by the balance models is that the mean and median are each located between the maximum and minimum observed values. And the workings of both models reinforce the formula or procedure needed to obtain numerical values for their respective measure of center.

One difference between the physical models, however, is that the pulley and loop demonstration actually automatically finds a median (hence the first word of this article’s title), whereas trial and error is needed to find the point of balance in the seesaw demonstration. Another difference we have identified is the role of mass. Suppose a teacher tries the mean balancing experiment by putting weights at various places on a yardstick. For the balance point to occur at the mean, the wood at the ends of the leftmost and rightmost weights must be equal in length, a major limitation on what datasets can be explored. The students’ textbook ([Perkowski and Perkowski 2007](#)) acknowledges this pitfall:

“Imagine the ‘x’s’ as weights on a long ‘weightless’ stick...If you used your finger as a fulcrum, the point on the stick at which it balanced would be the mean. This is unfortunately tricky to do in real life, as any type of stick we could use would have some mass (and weight) and throw our mean off a bit” (p. 51).

This is why it takes a loop to find the median, because if you just draped a string (with clamps attached in various places) not tied in a loop over a pulley, the weight of the string (assuming the string did not fall off the pulley) could cause the balance point to be different than the median. The mass of a loop, however, does not affect the loop’s balance point. The pulley-and-loop activity gives students another way to make their understanding concrete and gives instructors another way to make a basic statistics lesson an interactive, hands-on experience.

Appendix A: In-Class Assessment Protocol

Dataset	Assessment		
#1	Instructor Prompts <i>Based on the sample size of the data set, how many clips will we need?</i> [Place clips as described in the above the first row of Table 3] <i>Are the values already sorted? Why?</i>	Student Engagement Ask students questions to verify that they understand how to "read" the number loop. Ask a student to place or guide your placement of the clips to correspond to the observations listed.	Targeted Observations * There is an odd number of values, equally spaced. * Because the loop is formed from a number line, the values are already sorted.
	<i>When I let go of the string so the loop is free to move, what value will come to rest at the bottom of the loop?</i>	Students predict using choices: A) 24 (9 votes) B) something else (1 vote)	The number 24 quickly moves to the bottom of the loop.
	<i>Why do you think it was 24? Is 24 a particular measure of center?</i>	Discussion	The value 24 is actually the value of several measures of center in this case (mean, median, and midrange) and so more exploration is needed to make a conjecture.
	<i>Now, where should I put the fulcrum to make this "platform model" balance?</i>	Discussion	The midpoint value 24 is the only point on the stick where it balances.
#2	<i>When I let go of the string so that the loop is free to move, will the loop move to a new position?</i>	Students predict using choices: A) it'll stay (bottom of loop stays at 24) (9 votes) B) YES, loop will move in the direction where that outlier backs away from the wheel (0 votes) C) YES, loop will move in the direction where that outlier moves closer to the wheel (1 vote)	Making the minimum value into an outlier did not change this measure of center, so it cannot be the midrange, and it cannot be the mean. That suggests it may be the median, and indeed the value at the bottom is in the middle position of the ordered dataset.
	<i>As a followup, let's change the outlier's position from 10 to anything else between 7 and 16. When I let go of the string so that the loop is free to move, will the loop adjust to a new position?</i> [after this question is resolved, move the clip back to 10]	Students predict: A) it'll stay (10 votes) C) it'll change (0 votes)	Still no change: the exact position of that outlier does not matter. What does matter is that the middle position divides the data set into two equal sized groups, one on each side of the loop, so the first author created and shared this mnemonic jingle for students: "Each side of the loop's an equal-sized group!"
	<i>Now will the platform (model of the mean) still balance at 24?</i>	Students predict using choices: B) it will balance (0 votes) C) it will crater (10 votes)	No, it does not, because the mean is a measure of center that IS affected by an outlier. [By the way, the stick may not quite balance at the updated mean of 22 either because of

		<p>the asymmetry of the mass of the stick to either side. This is a limitation of the physical model of the mean that will be discussed later.]</p> <p>Note that while the exact position of the outlier did not affect the median, knowing all values exactly is obviously necessary to know the mean (i.e., where a weightless platform would balance).</p>	
#3	<p><i>When I let go of the string so that the loop is free to move, will a new value for the median be at the bottom of the loop?</i></p>	<p>Students predict:</p> <p>A) It'll stay (median will remain 24) (8 votes)</p> <p>B) Median gets a <u>boost</u> by 2 units (2 votes)</p> <p>D) median <u>declines</u> by 2 units (0 votes)</p> <p>C) <u>changes</u> to some other value (0 votes)</p>	<p>Yes, median increased by 2, from 24 to 26.</p>
	<p><i>What did adding 2 to all values do to the mean (which had been 22)? Will the platform balance now?</i></p>	Discussion	<p>The mean increased by 2 as well, from 22 to 24, and 24 is the (only) number for the mean where this platform balances.</p>
#4	<p><i>When I let go of the string so that the loop is free to move, will a new value for the median be at the bottom of the loop?</i></p>	<p>Students predict using choices:</p> <p>A) It'll stay (at 26) (0 votes)</p> <p>B) Median value gets a <u>boost</u> to 30 (5 votes)</p> <p>C) <u>Changes</u> to some other value (1 vote)</p> <p>D) Median value <u>declines</u> to 24 (3 votes)</p>	<p>It does not matter that 24 happens to be a repeated value – 24 is still in the middle position of all values of the data.</p>
	<p><i>Using the platform model, do we expect the mean to change?</i></p>	<p>A) It'll <u>stay</u> (2 votes)</p> <p>C) It'll <u>change</u> (8 votes)</p>	<p>The mean is where positive deviations and negative deviations sum to 0. Thus, the mean should not change (it's still 24, so stick still balances).</p> <p>Note that mean and median can be equal (both are 24), even though the distribution seems skewed.</p> <p>Also note that mean and median need not equal the halfway point between the maximum and minimum values; this halfway point is called the <i>midrange</i>, which equals $(12+30)/2 = 21$ in this case. Explicitly noting this is</p>

			helpful because it reminds students of the importance of precise language, lest “pick the middle point” get interpreted as the halfway point instead of the observation in the middle position. Confusion by “midpointer” students has been noted (e.g., Russell and Mokros 1990).
#5	<p><i>In the previous data set, some of you might not have been sure whether 24 was the median because it was the middle of 3 distinct values or because it was the middle of all 5 values.</i></p> <p><i>We will be forced to be clear about this idea in this updated dataset with 7 values, of which 5 are different.</i></p> <p><i>If we go by the middle position of all 7 values, the median will equal what? (24)</i></p> <p><i>But if we go by the middle position of the 5 different values, the median would equal what? (12)</i></p> <p><i>When I let go of the string so that the loop is free to move, will a new value for the median be at the bottom of the loop?</i></p>	<p>A) median stays at 24 (8 votes)</p> <p>B) median changes to 12 (2 votes)</p>	The value of a median in a dataset with repeated values can require taking into account the frequencies of the observations, as this dataset showed.
#6	<p><i>When I let go of the string so that the loop is free to move, will a new value be at the bottom of the loop?</i></p> <p><i>Is the mean affected by sample size being even or odd?</i></p>	<p>A) It'll stay (at 24) (9 votes)</p> <p>B) Yes, it will be something different from 24 (1 vote)</p>	When the two middle positions in an even number of observations have the same value, that common value is the median.
#7	<p><i>When I let go of the string so that the loop is free to move, will a new value be at the bottom of the loop?</i></p>	<p>A) It'll stay (at 24) (0 votes)</p> <p>B) Yes, median becomes 27 to <u>bisect</u> distance between 24 and 30 (9 votes)</p> <p>C) Median changes, but maybe not to 27 (1 vote)</p>	Unlike the previous experiments where the loop moved promptly to a new position when it had a move to make, this time the loop is not so decisive. The reason is because each side of the loop has an equal-sized group as long as the value at the bottom

of the loop is between 24 and 30, but it doesn't have to be halfway between 24 and 30, it can be ANY value (strictly) between 24 and 30. To show this, we can gently manually nudge the loop to values like 25, 26, 28, 29 and the loop will stay in place, in equilibrium.

This demonstrates that a dataset with an even number of observations can have a median that is not one of the values in the data set.

Appendix B: Pre-post Test Questions about the Median

1. The table below displays the number of student comments made for one class period with her eight-student class. The teacher wants to summarize this data by computing the typical number of comments made that day. Explain how reporting the mean number of comments could be misleading.

Student Initials								
	A.A.	R.F.	A.G.	J.G.	C.K.	N.K.	J.L.	A.W.
Number of comments	0	5	2	22	3	2	1	2

2. A test was given to 60 individuals. Half answered 80% of the questions correctly, the other half of the class answered 90% correctly. With regard to the mean and median of the class scores, which of the following statements is correct?
 a. mean > median b. mean = median c. mean < median.

Explain.

3. A set of data are put in numerical order, and a statistic is calculated that divides the data set into two equal-sized groups. Which of the following statistics was computed?
 a. mean b. range c. mode d. median

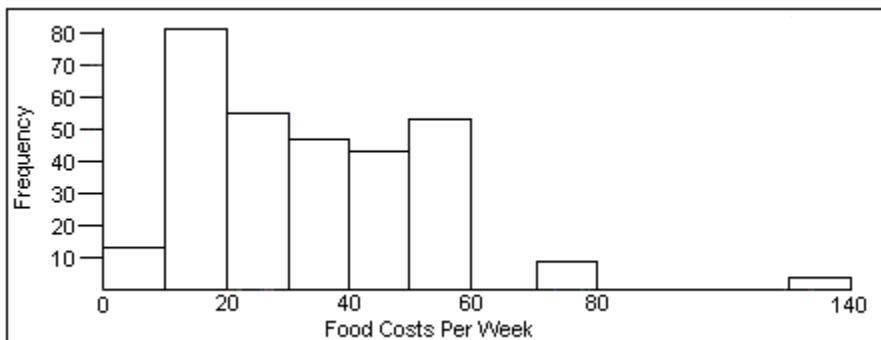
4. A test was given to 60 individuals. One third answered 80% of the questions correctly, the other two thirds of the class answered 90% correctly. With regard to the mean and median of the class scores, which of the following statements is correct?
 a. mean > median b. mean = median c. mean < median.

Explain.

5. Here are the number of hours that seven statistics students studied for this exam: 3, 5, 11, 6, 4, 2, 4. We could say that the median number of study hours is:
 a. 4 b. 4.5 c. 5 d. 6 e. 11

6. Here are the number of hours that seven statistics students studied for this exam: 5, 11, 6, 4, 2, 4. We could say that the median number of study hours is:
 a. 4 b. 4.5 c. 5 d. 6 e. 11

7. This is a distribution of how much money was spent per week for a random sample of college students. The following statistics were calculated: mean = \$31.52; median = \$30.00; range = \$132.50.

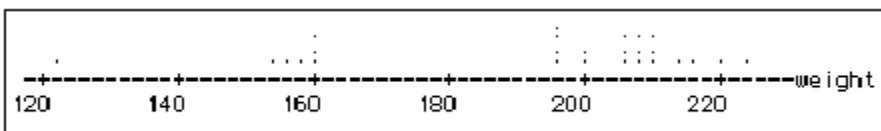


The median food cost tells you that a majority of students spend about \$30 each week on food.

- Agree, the median is an average and that is what an average tells you.
- Agree, \$30 is representative of the data.
- Disagree, a majority of students spend more than \$30.
- Disagree, the median tells you only that 50% spent less than \$30 and 50% spent more.

Items 8 and 9 refer to the following situation:

The following dotplot displays the distribution of weights of the members of the 1996 U.S. Men's Olympic Rowing Team:



- Estimate the value of the median of the distribution as accurately as you can from this plot.
- Would the mean be greater than or less than the median for these data? Explain briefly.
- NCAA wrestler weights are summarized in the table below. However, when determining these weights it was discovered that the scale topped out at 250 lbs. so that the exact weight of any player 250 lbs. or more could not be determined. The following is a sample of weights from a particular team:

Weight	Frequency
175	4
200	8
225	15
250+	9

Which can we say about the median?

- It cannot be determined
- The median is 175
- The median is 200
- The median is 225
- The median is 250
- The median can be determined, but is not one of the choices

Acknowledgments

The authors express appreciation to department lecturer (with a master's degree in statistics) Francis Biney and (mathematics education master's student) Heidi Heinrichs for independently rating student responses on the pre/post-test's open-ended questions. We also thank Art Duval for a helpful hallway conversation, journal editors and referees for excellent feedback, the class students for their participation, Judah Lesser for the photography ([Figures 1](#) and [2](#)), Leo Montañez for the videography, and Mark Lynch for his encouragement to take his initial inspiration to the next level.

References

American Statistical Association (2005), *Guidelines for Assessment and Instruction in Statistics Education: College Report*. Alexandria, VA: American Statistical Association.
<http://www.amstat.org/education/gaise/>

American Statistical Association (2007), *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*. Alexandria, VA: ASA.

Bock, D. E., Velleman, P. F., and DeVeaux, R. D. (2007), *Stats: Modeling the World* (2nd ed.). Boston: Pearson.

College Board (2013), *2013 College-Bound Seniors Total Group Profile Report*. New York: College Board.
<http://media.collegeboard.com/digitalServices/pdf/research/2013/TotalGroup-2013.pdf>

Cooper, L. L. and Shore, F. S. (2008), "Students' Misconceptions in Interpreting Center and Variability of Data Represented in Histograms and Stem-and-leaf Plots," *Journal of Statistics Education*, 16(2), 1-13. <http://www.amstat.org/publications/jse/v16n2/cooper.pdf>

Doane, D. P. and Tracy, R. L. (2000), "Using Beam and Fulcrum Displays to Explore Data," *The American Statistician*, 54(4), 289-290.

Falk, R. and Bar-Hillel, M. (1980), "Magic Possibilities of the Weighted Average," *Mathematics Magazine*, 53(2), 106-107.

Frankfort-Nachmias, C. and Leon-Guerrero, A. (2000), *Social Statistics for a Diverse Society* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Friel, S. N. and Bright, G. W. (1998), "Teach-Stat: A Model for Professional Development in Data Analysis and Statistics for Teachers K-6." In Susanne P. Lajoie (Ed.), *Reflections on Statistics: Learning, Teaching, and Assessment in Grades K-12* (pp. 89-117). Mahwah, NJ: Lawrence Erlbaum Associates.

Garfield, J. (1993), “Teaching Statistics Using Small-Group Cooperative Learning,” *Journal of Statistics Education*, 1(1). <http://www.amstat.org/publications/jse/v1n1/garfield.html>

Garfield, J. B. and Gal, I. (1999), “Assessment and Statistics Education: Current Challenges and Directions,” *International Statistical Review*, 67(1), 1-12.

Gfeller, M. K., Niess, M. L., and Lederman, N. G. (1999), “Preservice Teachers’ Use of Multiple Representations in Solving Arithmetic Mean Problems,” *School Science and Mathematics*, 99(5), 250- 257.

Groth, R. E. and Bergner, J. A. (2006), “Preservice Elementary Teachers’ Conceptual and Procedural Knowledge of Mean, Median, and Mode,” *Mathematical Thinking and Learning*, 8(1), 37-63.

Han, J., Kamber, M., and Pei, J. (2012), *Data Mining: Concepts and Techniques* (3rd ed.). Waltham, MA: Morgan Kaufmann Publishers.

Hanley, J. A., Joseph, L., Platt, R. W., Chung, M. K., and Bélisle, P. (2001), “Visualizing the Median as the Minimum-Deviation Location,” *The American Statistician*, 55(1), 1-3.

Hardiman, P. T., Pollatsek, A., and Well, A. D. (1986), “Learning to Understand the Balance Beam,” *Cognition and Instruction*, 3(1), 63-86.

Hardiman, P. T., Well, A. D., and Pollatsek, A. (1984), “Usefulness of a Balance Model in Understanding the Mean,” *Journal of Educational Psychology*, 76(5), 793-801.

Holmes, P. (2003), “From Sharing to Balancing – No Mean Feat!,” *Teaching Statistics*, 25(2), 36-39.

Hudson, R. A. (2012/2013), “Finding Balance at the Elusive Mean,” *Mathematics Teaching in the Middle School*, 18(5), 301-306.

Jacobbe, T. (2012), “Elementary School Teachers’ Understanding of the Mean and Median,” *International Journal of Science and Mathematics Education*, 10(5), 1143-1161.

Leavy, A. M., Friel, S. N., and Mamer, J. D. (2009), “It’s a Fird! Can You Compute a Median of Categorical Data?,” *Mathematics Teaching in the Middle School*, 14(6), 344-351.

Lesser, L. (1999), “The ‘Ys’ and ‘Why Nots’ of Line of Best Fit,” *Teaching Statistics*, 21(2), 54-55.

Lesser, L. (2011a), “On the Use of Mnemonics for Teaching Statistics,” *Model Assisted Statistics and Applications*, 6(2), 151-160.

Lesser, L. (2011b), “Low-Tech, Low-Cost, High-Gain, Real-Time Assessment? It’s all in the cards, easy as ABCD!,” *Texas Mathematics Teacher*, 58(2), 18-22.

http://www.tctmonline.org/TMT_archive/TMT_2010s/TMT_2011_Fall.pdf

Lesser, L. (2011c), Making statistics memorable: New mnemonics and motivations. *Proceedings of the 2011 Joint Statistical Meetings, Section on Statistical Education* (pp. 1118-1124).

<http://www.statlit.org/pdf/2011Lesser-JSM.pdf> or

https://www.amstat.org/membersonly/proceedings/2011/papers/300805_65461.pdf

Lesser, L. (2012), “Further Comments and Conceptualizations Regarding the Mean,” *Journal of Statistics Education*, 19(3), 1-3.

http://www.amstat.org/publications/jse/v19n3/lesser_letter.pdf

Lesser, L. (2013), “One ‘Mean’ Class,” *Statistics Teacher Network*, no. 81, p. 20.

<http://www.amstat.org/education/stn/pdfs/STN81.pdf>

Lesser, L. and Winsor, M. (2009), “English Language Learners in Introductory Statistics: Lessons Learned from an Exploratory Case Study of Two Pre-service Teachers,” *Statistics Education Research Journal*, 8(2), 5-32.

[http://iase-web.org/documents/SERJ/SERJ8\(2\)_Lesser_Winsor.pdf](http://iase-web.org/documents/SERJ/SERJ8(2)_Lesser_Winsor.pdf)

Lindgren, B. W. (1976), *Statistical Theory* (3rd ed.). New York: Macmillan.

Lynch, M. (2009), “Teaching Tip: The Median is a Balance Point,” *The College Mathematics Journal*, 40(4), 292.

Mallows, C. (1991), “Another Comment on O’Cinneide,” *The American Statistician*, 45(3), 257.

Martin, M. A. (2003), “‘It’s Like... You Know’: The Use of Analogies and Heuristics in Teaching Introductory Statistical Methods,” *Journal of Statistics Education*, 11(2).

<http://www.amstat.org/publications/jse/v11n2/martin.html>

Mayén, S. and Díaz, C. (2010), “Is Median an Easy Concept? Semiotic Analysis of an Open-ended Task.” In C. Reading (Ed.), *Data and Context in Statistics Education: Towards an Evidence-based Society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

http://iase-web.org/documents/papers/icots8/ICOTS8_C265_MAYEN.pdf

National Council of Teachers of Mathematics (2000), *Principles and Standards for School Mathematics*. Reston, VA: Author.

National Governors Association and Council of Chief State School Officers (2010), *Common Core State Standards (College- and Career- Readiness Standards and K-12 Standards in English Language Arts and Math)*. Washington, DC: National Governors Association Center for Best Practices and Council of Chief State School Officers. <http://www.corestandards.org>

Neumann, D. L., Hood, M., and Neumann, M. M. (2013), "Using Real-life Data when Teaching Statistics: Student Perceptions of this Strategy in an Introductory Statistics Course," *Statistics Education Research Journal*, 12(2), 59-70.
[iase-web.org/documents/SERJ/SERJ12\(2\)_Neumann.pdf](http://iase-web.org/documents/SERJ/SERJ12(2)_Neumann.pdf)

O'Dell, R. S. (2012), "The Mean as Balance Point," *Mathematics Teaching in the Middle School*, 18(3), 148-155.

Penafiel, A. F. and White, A. L. (1989), "SSMILES: Exploration of the Mean as a Balance Point – Grades 6-9," *School Science and Mathematics*, 89(3), 251-258.

Perkowski, D. A. and Perkowski, M. (2007), *Data and Probability Connections: Mathematics for Middle School Teachers*. Upper Saddle River, NJ: Pearson Prentice Hall.

Pollatsek, A., Lima, S. and Well, A. D. (1981), "Concept or Computation: Students' Understanding of the Mean," *Educational Studies in Mathematics*, 12(2), 191-204.

Roe, M., Shin, H., Ukkusuri, S., Blatt, A., Majka, K. et al. (2010), "The New York City Pedestrian Safety Study and Action Plan Technical Supplement," New York City Department of Transportation.

http://www.nyc.gov/html/dot/downloads/pdf/nyc_ped_safety_study_action_plan_technical_supplement.pdf

Russell, S. J. and Mokros, J. R. (1990), "What's Typical? Children's and Teachers' Ideas about Average." In David Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics*, Vol. 1 (pp. 307-313). Voorburg, The Netherlands: International Statistical Institute. <http://iase-web.org/documents/papers/icots3/BOOK1/A9-3.pdf>

Schwertman, N. C., Gilks, A. J., and Cameron, J. I. (1990), "A Simple Noncalculus Proof that the Median Minimizes the Sum of the Absolute Deviations," *The American Statistician*, 44(1), 38.

Shrout, P. E. and Fleiss, J. L. (1979), "Intraclass Correlation: Uses in Assessing Rater Reliability," *Psychological Bulletin*, 86(2), 420-428.

Stigler, S. M. (1977), "Fractional Order Statistics, with Applications," *Journal of the American Statistical Association*, 72(359), 544-550.

Strauss, S. B. and Bichler, E. (1988), "The Development of Children's Concepts of the Arithmetic Average," *Journal for Research in Mathematics Education*, 19(1), 64-80.

Székely, G. J. (1997), "Problem Corner," *Chance*, 10(4), 24-25.

Watier, N. N., Lamontagne, C., and Chartier, S. (2011), "What Does the Mean Mean?," *Journal of Statistics Education*, 19(2), 1-20.

<http://www.amstat.org/publications/jse/v19n2/watier.pdf>

White, A. L. and Berlin, D. F. (1989), "Fulcrum and Mean: Concepts of Balance – High School/College," *School Science and Mathematics*, 89(4), 335-342.

Zawojewski, J. S. and Shaughnessy, J. M. (2000a), "Data and Chance." In E. A. Silver and P. A. Kenney (Eds.), *Results from the seventh annual assessment of the National Assessment of Educational Progress* (pp. 235-268). Reston, VA: National Council of Teachers of Mathematics.

Zawojewski, J. S. and Shaughnessy, J. M. (2000b), "Mean and Median: Are They Really So Easy?," *Mathematics Teaching in the Middle School*, 5(7), 436-440.

Lawrence M. Lesser
Department of Mathematical Sciences
The University of Texas at El Paso
500 W. University Avenue
El Paso, TX 79968
Lesser@utep.edu

Amy E. Wagler
Department of Mathematical Sciences
The University of Texas at El Paso
500 W. University Avenue
El Paso, TX 79968
awagler2@utep.edu

Prosper Abormegah
Department of Mathematical Sciences
The University of Texas at El Paso
500 W. University Avenue
El Paso, TX 79968
pabormegah@utep.edu
