# The Psycho-Social and Environmental Variables Influencing Institutional Data Integrity

Don Turk, Ph.D.
Director, Institutional Research
Colorado Mountain College

Meeta Goel, Ph.D.
Dean, Institutional Effectiveness
Colorado Mountain College

Kendra Carpenter
Research Analyst
Colorado Mountain College

Barb Johnson
Research Analyst
Colorado Mountain College

Veneeya Kinion
Research Analyst
Colorado Mountain College

Jonathan Hansen
Research Analyst Assistant
Colorado Mountain College

## ABSTRACT

The present paper is an examination of the self-reported variables associated with data entry errors at a specific institution, relative to one specific database. Following a review of literature covering studies on data integrity, the method and results of an empirical study are described that demonstrates the discrepancy between objective error counts and subjective self-reported error tendencies and other related variables. End users of the Datatel database at a community college in Colorado (N = 42) self-reported on their: physical work conditions, feelings about using the database, tendency towards distraction, ease of using the database, concern about data integrity, and degree of being error prone. The results for these measures indicate a relatively strong tendency towards accurate data entry at the institution. However, objective error counts for a seven-month period of time demonstrate contrasting tendencies. The discrepancies between these data are attributed to self-serving biases and inaccurate self-perceptions on the part of end-users. Practical recommendations extending from the results are provided for improving institutional data integrity.

# The Psycho-Social and Environmental Variables Influencing Institutional Data Integrity

## Introduction

In current times, the ubiquity of computer based data input systems demands that significant trust be placed in the accuracy of data which fill the plethora of data banks utilized throughout higher education. [1] Much (if not most) of the time such trust seems warranted. Data entry personnel are usually trained, retrained, and reminded extensively about the importance of data integrity. Though it is safe to assume that most data entry personnel are at least vaguely conscientious about entering accurate data, even the best employees make mistakes, which may go undetected sometimes for months or even years.

Of course, for most databases, there may appear to be relatively few data entry errors within any given institution, at least when compared to those data which are entered correctly. And, small numbers of errors can yield perceptions that the topic of data integrity is rather irrelevant compared with other loftier institutional aims such as student satisfaction, learning, and access. Ironically, such an argument is itself in err and ignores the compounding effect that small numbers of errors have on the data reporting agencies of an institution such as Institutional Research or Institutional Effectiveness. For example, one erroneous student headcount entry can have virtually exponential effects as such data are reported to various domains of an institution, which in turn generate their own reports on the basis of the faulty figure(s). Changes in policy, funding cuts, and a variety of other negative consequences within an institution can be generated on

the basis of erroneous data. Hence, in the present research, we assume that data integrity is as fundamental to the educational enterprise as sound pedagogical practice and strong student development infrastructure.

Due to the importance of this topic for the institution with which the authors are affiliated, it has become imperative to develop a thorough understanding of human error as it relates to data integrity. The present paper is an exploratory examination of the self-reported variables associated with data entry errors relative to one specific database. Following a review of literature covering studies on human error and data integrity, the method and results of an empirical study are described that demonstrates the discrepancy between objective data entry error counts and subjective self-reported data entry error tendencies. The primary foci of the study are the discrepancies between objective error tallies and the self-reported tendencies from participants. Practical recommendations extending from the results are provided for improving institutional data integrity.

## REVIEW OF RELEVANT LITERATURE

As we progress through the information age most of us are becoming more and more sensitive to occasional inaccuracies embedded within the informative and persuasive messages proliferating our day-to-day lives. Of course, especially in the realm of advertising, some of those inaccuracies are perhaps intentional "truth stretching." However, much of the fallacious information we are exposed to is a product of human or technological error, especially in the arena of news reporting. Given the sea of information within

which we swim, the statistical probability of encountering error is rather great these days, but how much of a relevant impact does it really have for most of us in our personal and occupational lives?

The consequences of human error relative to data entry has been noted by the IT Policy Compliance Group who recently reported that a fifth of organizations suffer "22 or more sensitive data losses a year" and that user error is responsible for half of such losses (Shifrin, 2007). According to the report, an additional 25% of data losses are due to deliberate or accidental policy violations on the part of employees. The consequences of such losses are largely financial for the organizations included in Shifrin (2007), but it is important to note that data entry errors result in more than mere losses of revenue when it comes to the educational enterprise.

For example, in 2006, the state of Illinois had a lengthy delay in the release of test-scores for their Standards Achievement Test due (in large part) to data-entry errors. Close to 11,000 tests had missing or incorrect demographic data (McNeil & Hoff, 2006). The Standards Achievement Test scores are assessed prior to allocating resources to students on the basis of the No Child Left Behind Act. The consequences of the delayed release of the scores were that many students "missed out on free tutoring or the ability to transfer out of schools." Ultimately the state had to make "educated guesses" about which of their 421,000 students would be entitled to receive free tutoring.

Despite its importance, the topic of data integrity within educational institutions has not been widely addressed by researchers in peer review

contexts. When data integrity is the focus of academic research, it is typically in the context of human error probability estimations for industrial and medical settings, where errors can result in loss of human life (e.g., Khan, et al., 2006; Koren et al., 2007; Park & Lee, 2008). Khan et al. (2006) focused their research efforts on the burgeoning offshore industry in the North Atlantic region of the United States, where the fatal consequences of accidents seriously warrants the development of means for quantifying human error. Khan et al.'s (2006) use of human error probability indices (HEPI) was aimed at limiting the "opportunities for human error occurrence" and mitigating the "results of such errors through changes in training, design, safety systems and procedures, resulting in a more error tolerant design and operation (p. 313). Essentially, HEPI is a technique that calculates a probability of human error by dividing the number of errors that occur in a given task by the number of opportunities available for committing an error.

Park & Lee (2008) offered a refinement to the HEPI due to their belief that the HEPI process involves too great a degree of subjectivity and forces "heavy reliance on the judgments of experts in the field" (p. 578). Park & Lee (2008) tested a HEPI estimation technique that quantifies subject judgments and confirms consistency of collected data.

Both of the studies described above (Khan et al., 2006; Park & Lee, 2008) approach human error as a problematic inevitability and have designed methods to assist in predicting and controlling error as effectively as possible. This approach is atypically refined for the small number of studies examining human error in various contexts. Most research in this area is quite exploratory and

descriptive rather than predictive or explanatory. One (of few) additional exception to this trend is Koren et al. (2007), who attempted to develop an explanatory model of errors among nursing students that included two dimensions; (1) Error Prone Persons and (2) Error Prone Environments. The authors conclude that Error Prone Environments tend to be the most powerful influencer of errors among nursing students in clinical settings. Koren et al.'s (2007) explanation of error as a bi-dimensional phenomena (i.e., one dimension internal to the person committing the error and a second dimension being external to the person or in the person's environment) is a relatively well-documented approach to examining the causes of error (e.g., see Dekker, 2005; Dekker, 2006; Whittingham, 2004). In the next section of the review, we return to this bi-dimensional approach and develop it as the primary theoretical framework of the present study.

Dekker (2006) has provided a thorough overview of what he calls the "old view" of understanding human error. This view constructs errors within organizational, institutional and industrial contexts as being primarily attributable to people. In simple terms this "bad apple" view implies that unreliable people undermine systems where data are entered, interpreted, or otherwise acted upon. In the "bad apple" perspective, errors are attributed to the internal characteristics (personality, attitudes, values, beliefs, etc.) of the people who make errors. Dekker explains how this "bad apple" view results in institutional efforts to keep people in line by implementing greater standards of accountability and more punitive reprimands for making errors. Dekker goes on to describe how this "old

view" of human error largely ignores the systemic components that influence humans to make errors. In Dekker's view, the "bad apple" perspective does not fully account for the psychological processes in play as persons attempt to meet and sometimes reconcile conflicting goals or the systemic demands that are part and parcel of their day-to-day work processes. The new view of human error attributes error to the systemic and socio-cultural contexts within which humans operate as they do their work.

Whittingham (2004) supports this "systems not people" approach and develops an extensive overview of systemic error analysis demonstrating that the vast majority of errors (within industrial contexts, at least) are preventable because they result from defective systems within the organization. He explains how the "bad apple" view (articulated by Dekker, 2006) has contributed to a "blame culture" that emphasizes internal attributions whilst ignoring the error-inducing systems that people work within. While Dekker (2006) and Whittingham (2004) offer a useful reminder that errors are often not due solely to unreliable humans, it is worth noting that Dekker's and Whittingham's analyses focus largely on industrial contexts where the systems that humans are operating within are very complex. Such systems, (due to sheer statistical probability) are more error prone, if for no other reason than the opportunities for error are greater. For example, a worker functioning in a complex industrial context like air traffic control has a vast array of tasks to complete and decisions to make in (often) a small amount of time, with life and death consequences attached to those tasks and decisions.

The context of the present study is a far cry from the richly complex work environments that Dekker's and Whittingham's analyses focus on. In short, data entry into a database is perhaps not so complex as to assume no credulity for the "bad apple" view. Hence, in the present study we assume a more bi-dimensional approach to understanding data entry errors and assume that data-entry errors can be a consequence of features both external and internal to humans. It is our view that sometimes people do make errors due to carelessness (or other internally attributable causes) in spite of a system that may be designed to make errors not as likely.

Due to the fact that most human error analysts focus their attention on complex industrial contexts, this bi-dimensional approach (if it is shared or articulated to any extent by other researchers) is not readily available in the academic literature. There are therefore, no measures or taxonomies of error internality or error externality that are available for our examination and testing. Because of this, the measures in the present study (while examined with standard hypothetico-deductive means) were constructed inductively through the use of institutionally relevant anecdotes, group conversations, and reflections on our experiences with data entry. We assume based on these inductive techniques that data entry errors are likely attributable to the following variables:

- The tendency of some persons to be error prone or to lack the conscientiousness required for accurate data entry.
- The physical environments within which data entry occurs

- Affect and cognitions that data entry personnel hold towards the database and data entry itself

An additional gap in the literature on human error analysis is the lack of self-report data detailing the internal thoughts and feelings of the persons that are often the foci of the analyses. In the present study we have therefore opted to gather such self-report data and pose the following research questions as the basis of the examination:

> *How do end-users of a specific database view their own and others' data-entry errors? Do they rate their work environments in ways that appear conducive to accurate data entry? How do data entry personnel account for their data entry errors? Do they attribute their data entry errors more to themselves, to the database, or to the environments within which they work?*

## METHOD

### Participants

Involvement in this study was limited to persons with institutionally approved access to the Datatel [2] database and who have input data into the database at least once within the past four months (i.e., since October of 2007). Hence, the sampling frame for the study was 292 full or part-time staff. Approximately 90 of those staff (~ 30%) were Datatel users whose primary work responsibilities involve entering data into Datatel on a daily basis. The final sample for this study included 42 participants (mean age = 45; SD = 8.80; female = 35, male = 5, sex not reported = 2) most of which (n = 23, 55%) reported being employed with the institution for "5 + years." Four participants (10 %) reported being employed at the institution between 1-6 months. An additional ten percent

reported being employed there between one and three years. Six participants (14%) reported being employed there between three and five years.

Measurement

As noted previously, prior to conducting the present study, there were no available measurements for gauging the self-reported data entry errors and attributions of data-entry personnel. Hence, concerted effort was placed on developing a measure that would reliably and validly represent the range of relevant dimensions for the study. To assist with this task, a team of five persons all affiliated with the Department of Institutional Research met on several occasions to discuss the form and content of the measurement instrument. The substance of these discussions surrounded the plethora of internal and external factors that could possibly bear upon the accuracy of data entry for personnel. A preliminary set of dimensions was crafted and shared with a committee primarily responsible for managing the infrastructural support and policies surrounding Datatel.

In addition to the demographic items mentioned above, the final measurement instrument (see Appendix 1) included a set of 57 six-point Likert items designed to gauge participants' thoughts and feelings about: (1) entering data into Datatel, (2) the environments within which they work, and (3) attributions about data entry errors. There was one open-ended item that invited participants to describe a circumstance in which they have knowingly entered data incorrectly into Datatel.

Procedures

Following approval from the Institutional Review Board, the measure was loaded online using standard web survey construction software. Persons in the sampling frame were solicited to be in the study via email. Willing participants signed and submitted a voluntary consent form prior to accessing the survey. All survey data were collected anonymously. The survey was made available online for approximately two weeks. Most participants completed the survey within the first week of its availability. On average, each participant took about 10-15 minutes to complete the survey. As a compensation for their time and energy, participants were offered a $10 gift certificate to one of the institution's four book/gift stores.

## RESULTS

### Scale Creation and Reliability Analyses

The validity of the measurement instrument used in this research was assessed at face level [3] and then corroborated with Cronbach Alpha reliability coefficients. Below are the results for the scales created from the measure using this technique. The means and standard deviations for each of the created scales are presented in the section of the results following the reliability analyses.

Temperature

There were five items on the survey that asked participants to rate how comfortable they are with the temperature in their work environments (items 3, 12, 18, 21, & 25). The results of the reliability analysis indicated that items 12, 18, 21,

& 25 consistently measured feelings about temperature comfort. The items above were used to compute the variable TEMPERATURE (Cronbach's Alpha = .80).

Feelings about Datatel

There were 13 items on the survey designed to gauge how positively or negatively participants feel about entering data into Datatel (items 1-2, 4, 13, 15-17, 19-20, 23-24, & 55-56). The results of the reliability analysis indicated that when considered together, almost all 13 of the items (excluding item #2) provided a consistent indication of participants' positive or negative feelings about entering data into Datatel. The 12 items above were used to compute the variable FEELINGS (Cronbach's Alpha = .89).

Distractedness

There were eight items on the survey designed to gauge how distracted participants are when entering data into Datatel (items 5-10, 14, & 27). The results of the reliability analysis indicated that excluding items 5 & 10 provided an acceptable level of measurement consistency for the scale. The six items noted above were used to compute the variable DISTRACTED (Cronbach's Alpha = .70).

Datatel Easy/Supported

There were seven items on the survey designed to gauge participants' perceptions of the degree of difficulty associated with Datatel and/or the level support/training they have received relative to Datatel (items 11, 22, 26, 29, 30, 37 & 47). The results of the reliability analysis indicated that excluding items 37

& 47 provided an acceptable level of measurement consistency for the scale. The five items of the scale were used to compute the variable EASY (Cronbach's Alpha = .74).

Concern for Data Integrity

There were 17 items in the survey designed to measure the level of concern that participants expressed about entering data accurately or data integrity in general (items 28, 32, 34, 39-44, 46, 48, 49-51, 53-54, & 57). The results for the reliability analysis indicated that excluding items 32, 43, 48, & 51 provided an acceptable level of measurement consistency for the scale. The 13 items of the scale were used to compute the variable CONSCIENTIOUS (Cronbach's Alpha = .71).

Tendency to Make Errors

There were six (quantitative) items in the survey designed to measure the frequency with which participants enter data incorrectly into the Datatel system (items 31, 35-36, 38, 45, & 52). The results of the reliability analysis indicated that excluding items 38 & 52 provided an acceptable level of measurement consistency for the scale. The four items of the scale were used to compute the variable ERRORPRONE (Cronbach's Alpha = .74).

## Aggregated Results for Scales

The scales created and described above were measured using a six-point Likert metric. As the data were entered, the following numeric values were coded for each participants response for items 1-57 in the questionnaire: "strongly disagree" = 1; "disagree" = 2; "somewhat disagree" = 3; "somewhat

agree" = 4; "agree" = 5; "strongly agree" = 6. The scales were coded so that higher numbers would be indicative of higher levels of agreement with the variable in question. The means and standard deviations for each of the scaled variables are presented below.

TEMPERATURE

The scale labeled "TEMPERATURE" was a measure gauging how comfortable participants perceive their work environment to be in terms of warmth or coolness. Where needed, the items in the scale were recoded so that higher numbers would be more indicative of greater levels of agreement that participants feel the workspace temperature is comfortable. In general, participants agree that their workspace is a comfortable temperature. The mean for the scale fell between "somewhat agree" and "agree" (n = 39; Mean = 4.47, SD = .76).

FEELINGS

The scale labeled "FEELINGS" was a measure gauging how positively participants feel about entering data into Datatel. Where needed, the items in the scale were recoded so that higher numbers would be more indicative of agreement that entering data into Datatel is a positive experience. For the most part, participants appear to think of the data entry process in Datatel as both a somewhat positive and somewhat negative experience. The mean for the FEELINGS scale fell between "somewhat disagree" and "somewhat agree" (n = 40, Mean = 3.87, SD = .76).

DISTRACTED

The DISTRACTED scale was a measure gauging how distracted participants feel when they are entering data into Datatel. Where needed the items in the scale were recoded so that higher numbers would be more indicative of agreement that the participant is distracted when entering data into Datatel. In general, participants appear to be somewhat focused when entering data. The mean for the DISTRACTED scale fell between "somewhat disagree" and "somewhat agree" (n = 41, Mean = 3.21, SD = .72).

EASY

The EASY scale was a measure gauging participants' perceptions that Datatel is easy to use and is supported effectively. Where needed the items in the scale were recoded so that higher numbers would be more indicative of higher levels of participant agreement that Datatel is easy to use and is supported sufficiently with training, etc. Overall, participants appear to express (at least minimal) agreement that Datatel is easy to use and is supported by the institution (n = 40, Mean = 4.00, SD .77).

CONSCIENTIOUS

The CONSCIENTIOUS scale was a measure gauging participants' perceptions that data entry accuracy is an important topic worthy of their concern and attention. Where needed the items in the scale were recoded so that higher numbers would be indicative of greater agreement among participants that concerns for data entry accuracy and integrity are important. In general participants tend to agree that data entry accuracy and integrity are important.

The mean for the scale fell between "somewhat agree" and "agree" (n = 37, Mean = 4.79, SD = .48).

ERROR PRONE

The scale labeled ERROR PRONE was used to measure participants' self-reported tendency to make data entry errors or to enter data into Datatel inaccurately. Where needed the items in the scale were recoded so that higher numbers would be indicative of greater participant agreement that they are error prone. For the most part, participants tend not to agree that they are error prone. The mean for the scale fell between "somewhat disagree" and "disagree" (n = 40, Mean = 2.55, SD = .72).

Level of Data-Entry Competence and Years Employed

Participants were asked to report on a scale of 1-10 (1 = totally incompetent, 10 = highly advanced level of competence) the level of competence they possess in correctly entering data into the Datatel system. The mean for the item was 7.88 (SD = 2.12). Clearly people view themselves as more than minimally competent. Participants were also asked about their data entry experience prior to working at CMC (1 = no experience, 10 = highly experienced). The mean was 6.10 (n = 41, SD = 3.16). Eleven participants (27%) rated themselves as 10 in both competence and experience.

Open-Ended Question

In addition to the quantitative items described above, participants were given the opportunity to describe a circumstance in which they had entered data into Datatel that they knew were incorrect (or later found out were incorrect). In

addition, the question asked for the respondent to include the reason they entered the data incorrectly and/or their response to finding out that the data had been entered incorrectly. This question generated 25 responses. Of those 25 responses, 32% (n = 8) admitted that they knowingly entered incorrect data while 48% (n = 12) indicated that they were later informed that the data they entered was incorrect.

As to the reasons why they had entered data incorrectly, whether knowingly or not, the participants' responses were grouped into five themes: "In a hurry/rushed," "Misread or given incorrect data," "Datatel causes mistakes," "New employee," and "Didn't know or understand." A final category labeled "Miscellaneous" was created to capture all statements that would not discreetly fit into the other categories. Nine respondents (36%) indicated that the reason why they entered incorrect data was that the data they were given was illegible or they were not given correct data to enter. This category included responses like "Ideally the issues have come about due to the fact that what is being entered is not readable so you have to do a little guess on a letter or number..." or "I do not have the correct data to enter."

Six respondents (24%) stated that they were not aware that they had entered incorrect data. This category included statements such as "I didn't know I was entering wrong information until after I printed out a report," and "I was not aware that there was a space for filing which type of telephone line was to be used..." Three respondents (12%) felt that the incorrect data was entered because they felt rushed and needed more time to enter data correctly. This

category included comments like "Entered incorrect information only because was in a hurry to register a student and did not have anyone around to check with," or "I entered a grade incorrectly. It was the last set of grades to enter and I was trying to rush to meet the deadline."

An additional three respondents (12%) felt that Datatel causes data to be incorrectly entered. This category included comments like "I find that the way Datatel corrects you causes more errors." Two respondents (8%) indicated the reason for the data being entered incorrectly was because they were new in their positions and included comments like "This is partly because I am new in the position and am still learning the process and I probably thought I was doing it correctly." Two responses were categorized as "Miscellaneous," which included written comments like "n/a." In addition, the participants were asked to indicate in their answer what their response was to finding out that the data they had entered had been entered incorrectly. Seventy-two percent stated that they corrected their errors. It is important to note that not one participant attributed their data entry errors to carelessness or a lack of conscientiousness on their part.

An "Objective" Count of Datatel Errors

Since July 7, 2007, the Department of Institutional Research at the institution has been attempting to track the errors found in specific domains of the Datatel database. Given the enormity of the Datatel database it was decided that tracking errors in the most departmentally relevant portions of the database would be an efficient starting point for this process. Hence, errors observed

when preparing the monthly Full Time Equivalency (FTE) report were allocated for inclusion in the present study. [4]

The FTE report is generated for various decision makers in the college to ascertain the theoretical amount of full-time credit hours comprised by the credit and non-credit student body. There are several places within the preparation process of this report where errors could feasibly be found in the Datatel database. To generate this report the database is first queried and the results loaded into statistical analysis software. A general observation of the results ensues including the running of syntax with edits. The initial results are scrutinized and then shared with appropriate campus entities. The errors that are observed are corrected (if possible) and a final report is prepared and disseminated throughout the college.

To track the errors as specifically as possible during the creation of the FTE report, a spreadsheet was created with columns allocated for the following: (1) the date the error was observed, (2) the report that was being generated when the error was observed, (3) the type of error observed (i.e., "entry" or "omission"), (4) the file in Datatel within which the error was observed (5) the field in the Datatel file within which the error was observed, (6) a general description of the error, (7) the supervisor most likely to be in charge of the end-user(s) responsible for the error, (8) the total number of errors found for the observation and (9) a section for general comments about the specific error(s) observed.

Using the tracking process described above, a total of 1,582 errors were observed in the Datatel database during the seven months of data collection

we've chose to include in the present report. Sixty-nine percent (n = 1,084) of the errors were of the "omission" type (i.e., fields left blank that should be filled with data). The remaining 498 errors (31%) were "entry" errors (i.e., an incorrect value was placed in the field).

## DISCUSSION

This study has been an exploratory attempt to understand a few of the psychosocial and environmental variables that might be associated with data entry errors at one institution involving one specific database. Quantitative data were gathered from 42 end-users regarding their feelings about the temperature of their workspace, their feelings about the database (in general, as well as its ease of use and supportedness), their degree of distractedness when entering data in the database, the level of conscientiousness they possess relative to data integrity, and their self-perceptions regarding tendencies towards making errors. Supplemental (though minimal) qualitative data was also gathered which asked participants to describe and explain a situation in which they've entered data incorrectly and how they responded to that error. Finally, an objective tally of errors was conducted for a very limited dimension of the database over the course of seven months. The notable discrepancy between the self-reported data and the objective error count reported in the results yields a multitude of different observations, conclusions and foci for discussion. These are included below, in the final sections of the manuscript.

Environmental Factors

Apparently the workspaces for end-users who enter data into the database are not perceived to be a factor associable with data entry errors. In general, participants at least mildly agree that their workspace is a comfortable temperature. Interestingly, the data for the one item that was deleted from the TEMPERATURE scale due to its alpha-decreasing outcomes (item number 3 – See Appendix 1) yielded results arguably in contrast with those found for the temperature scale. When presented with a statement about work conditions being "ideal," a very substantial portion (n = 19, 46%) either somewhat disagreed, solidly disagreed, or strongly disagreed with the statement. Of course, the TEMPERATURE scale assesses feelings about the warmth and coolness of the workspace and perceptions about physical/environmental work conditions include much more than thoughts about temperature. Unfortunately, the survey did not examine other physical variables in the workspace besides temperature, so it is impossible to ascertain whether these 46% who feel the workspace is not ideal are dissatisfied with the temperature or some other feature of their work environment. It may very well be the case that other non-temperature related features of the work environment (e.g., aesthetics of the workspace, lighting, noise levels, smell, etc.) exert a notable influence on data entry among some personnel. This issue should be examined in a follow-up study.

Feelings about the Database

Participants appear to think of the data entry process in Datatel as both a somewhat positive and (not surprisingly, given that they are entering data as part

of paid work) a somewhat negative experience. The mean for the FEELINGS scale fell between "somewhat disagree" and "somewhat agree." It was assumed that the level of positivity and/or negativity that end-users possess towards the data base might be at least somewhat illuminative of the number of errors observed in the error report. However, such does not appear to be the case, as most end-users who responded to this survey do not seem to possess any strong feelings at all about the database. It is possible that feeling somewhat neutral about Datatel is associated with being cavalier and/or careless about data entry accuracy, but this conclusion is not supported by the data indicating end-users perceive themselves to be quite conscientious and competent. In short, the feelings that participants expressed about Datatel do not suggest that data entry errors are a result of negative affect that end-users hold regarding the database.

Distractedness & Ease of Use

The same is likely true regarding the possibility that end-users are overly distracted while attempting to enter data into the system. In general, participants appear to be somewhat focused when entering data. The mean for the DISTRACTED scale fell between "somewhat disagree" and "somewhat agree. Overall, participants appear to express at least minimal agreement that Datatel is easy to use and is supported by the institution.

Conscientiousness & Tendency Toward Error

Perhaps the most perplexing are the results for the scales measuring respondents' conscientiousness about data integrity and their self-reported tendencies to make errors in the system. Given the large number of errors

observed in the error report it would seem reasonable that participants' level of concern for data integrity would be lower than observed or the self-reported tendency towards errors would be greater. In general, participants tend to agree that data entry accuracy and integrity are important and they tend not to agree that they themselves are error prone.

The Power of the Self-Serving Bias

The psychological literature is replete with concepts and theories that can assist in explaining the present data. Psychologists have long been aware that humans demonstrate a "self-serving bias" when it comes to formulating accounts and/or ratings of their own behavior, especially when the behavior has self-esteem related consequences (Tesser, 2000). Even though the data for this study were collected anonymously, it is reasonable to assume that the present measurement may have elicited a self-serving bias. The objective error report clearly demonstrates that errors are prolific but not one of the variables that we measured explains directly or indirectly how or why these errors abound. The possibility that people are much more error prone and a lot less conscientious than they report they are seems to be at least one reasonable explanation. This conclusion is corroborated by the data for the open-ended survey item. Not one participant attributed his or her data entry errors to carelessness. In virtually every case, errors were attributed to external causes. It is also important to note the possibility that the mildly negative feelings that participants report about Datatel could be linked with the errors, but there is no reasonable explanation for how this affect could account for such a large number of errors.

Limitations

An additional explanation and limitation in the present study is that there is little or no reason to be certain the people responsible for the errors included in the error report are the same people who participated in the present study. Though it is unlikely that individuals other than those who participated in the research made the errors, we are not able to be verify it with certainty because the participants self-selected for the study and all the data were gathered anonymously.

An additional limitation surrounds the nature of the Likert measure used in this research. It is well known that retrospective self-reporting is flawed, especially when it comes to self-esteem relevant variables (as noted above). For example, participants were asked to rate their tendency towards data entry errors under the presumption that they are knowledgeable of when they are in err. Many errors may be (and perhaps are likely to be) committed in the Datatel system without awareness. In fact, participants reported in the open-ended question that often they found out that their entries were in error only retrospectively.

Another possible interpretation and limitation of the present study is that the variables examined here were not those that could account for the data entry errors in the Datatel system. Due to resource limitations the measure was limited in length. Perhaps errors are due to all kinds of other variables we've not chosen to measure. As this issue is examined more thoroughly in the future, an

exhaustive list of all possible internal and external variables associated with data integrity should be explored.

Conclusion & Directions for Future Research

Not withstanding the limitations described above, the present study has still yielded some useful information and should be applied across the institution. One possible direction of application that could extend the present study is to disseminate the error report described herein and ask for participants to account for specific errors. Even if the source of the error was not willing to admit to the error (or they were unaware of having made the error), other Datatel end users could possibly provide an account based on their own reasoning or thought processes when entering the same or similar data. At the barest minimum, such an approach could possibly illuminate the huge discrepancy between self-reported error tendencies and the actual amount of errors that are being entered into the system. End users might leave such a situation aware of the ease with which data can be entered inaccurately and this might bolster even more conscientiousness.

A really useful research endeavor would be an actual study simulating the work environment within the Datatel system so that an account of errors can be developed that is more direct and does not involve retrospective self-reporting. The current study attempts to link error counts with self-reported variables that can only hypothetically be linked to the individual errors reported.

## End Notes

[1] This claim applies to those databases specific to institutions (e.g., Datasync, POISE, Datatel, etc.) as well as databases that are more broadly used by numerous academic institutions (e.g., IPEDS). Continual cross validation of data derived from such databases would require more time and energy than most institutional employees have available.

[2] The Datatel system, also called Colleague, allows students, faculty, and staff to both input and output information. The system is designed so that each file is user-limited based on criteria including employment status and placement in the institution. For example, students, through the WebPortal, are able to update emergency information, register for classes (using WebAdvisor), and also link to library resources as well as the Online courses (using WebCT). Faculty members are able to use the Datatel/Colleague system to gain access to the student information system and posting grades. Administrative staff members use the Datatel/ Colleague system to make changes, create records, and access student records. With Datatel, the Admissions Office tracks correspondence with applicants until they have made their deposit and are registered for their first semester. The Business Office uses Datatel to record tuition and housing payments. Academic Advising can use the system to record students' area of study. Residential Life can use the system to assign student room numbers.

[3] Obviously, measurement validity is an important consideration even for

exploratory research studies like the present one. Exploratory and confirmatory factor analysis would therefore be highly recommended as a way of empirically substantiating the validity of measured dimensions. However, given the small sample size for the present study and the nature of our research goals (i.e., primarily the data are to be used internally), we've opted to forgo a factor analysis for the measure and assess validity on the basis of whether the items gauging particular "factors" appear reasonable by common sense standards and meanings in the English language (i.e., "face validity"). We've corroborated this approach by providing Cronbach's Alpha reliability coefficients for all factors with the underlying presumption being that if the measured dimensions are at least indicating stability and consistency, then they might also be valid as well.

[4] It is important to note that the fields in Datatel necessary for completing the monthly FTE report represent a very small portion of the database. The errors observed while preparing the FTE report are likely to grandly underestimate the actual number of errors entered into the entire database.

Appendix 1 Measurement Instrument (Note: Format of actual survey was visually different than what appears below. The content is the same though).

PLEASE REMEMBER: ALL OF YOUR RESPONSES ON THIS SURVEY ARE COMPLETELY ANONYMOUS. NO ONE (NOT EVEN THE RESEARCHERS) WILL EVER KNOW YOUR INDIVIDUAL RESPONSES. YOUR HONEST AND SINCERE PARTICIPATION IS HIGHLY APPRECIATED.

Please answer the questions in the order they are presented. This survey software does not permit the use of the "back" arrow while participating in the survey.

I. The scales below are designed to assess your perceptions about entering data into the datatel system. For each item, please indicate your level of agreement or disagreement with each statement by clicking the circle most accurately reflecting your feelings.

Strongly Disagree Disagree Somewhat Disagree Somewhat Agree Agree Strongly Agree

1. I look forward to entering data into datatel.
2. When I am NOT at work I reflect upon how much I enjoy entering data into datatel.
3. The conditions within which I work are ideal for entering data into datatel.
4. Using datatel makes me feel angry.
5. I input data into datatel when I feel sick or ill.
6. My job requires that I input data into datatel while working on other tasks simultaneously.
7. When I input data into datatel, I am highly focused.
8. I concentrate on the task at hand when I am entering data into datatel.
9. I input data into datatel when I am tired.
10. When I input data into datatel, I worry about whether I'm doing it the right way.
11. I feel like I have been well trained to accurately input data into datatel.
12. The temperature in my workspace is too hot when I am entering data into datatel.
13. Entering data into datatel is intimidating to me.
14. My job requires that I rush when I enter data into datatel.
15. I feel bored when I enter data into datatel.
16. I like to enter data into datatel.
17. I feel frustrated when using datatel.
18. The temperature in my workspace is too cold when I am entering data into datatel.
19. I feel confused when I enter data into datatel.
20. Entering data into datatel makes me happy.
21. My workspace is the perfect temperature for entering data into datatel.
22. Support help for datatel is easily available.
23. I am distracted when I enter data into datatel.
24. I am under pressure when I enter data into datatel.
25. The temperature in my workspace is uncomfortable for me when I am entering data into datatel.
26. My computer makes it easy for me to enter data into datatel.
27. When working, I have enough time to enter data into datatel accurately.

28. The relaxed culture at CMC makes me unconcerned about my data entry errors in datatel.
29. It is very easy to make a mistake when entering data in datatel.
30. Entering data accurately into datatel is easy.
31. I never make mistakes in datatel.
32. People who make mistakes in datatel should be fired immediately.
33. My colleagues make more mistakes in datatel than I do.
34. I know my colleagues make mistakes in datatel, but I don't tell anyone about it.
35. I often make mistakes in datatel.
36. If my supervisor knew how many errors I make in datatel, she/he would be surprised.
37. Usually, the errors in datatel are due to datatel and not the people who enter data.
38. In my job, data errors would occur more frequently if I weren't as careful as I am.
39. Datatel errors have a way of correcting themselves over time.
40. If I make a mistake in datatel I know that someone else will correct it.
41. I have no concern about whether data is entered into datatel correctly.
42. I'm not paid enough to make sure that data is entered correctly into datatel.
43. If I notice data that have been entered incorrectly into datatel, I always correct it.
44. I have intentionally entered data incorrectly into datatel.
45. I hope no one ever finds out how many errors I make in datatel.
46. Entering data accurately into datatel is important.
47. If I knew more about how to use datatel, I wouldn't make as many errors when entering data.
48. I'd rather enter data incompletely than incorrectly.
49. Even the best employees often enter data incorrectly into datatel.
50. I don't have time to worry about other people's data entry mistakes in datatel.
51. I see no problem in sharing my datatel password with other CMC employees who need to use it to access datatel.
52. At times data in datatel is entered incorrectly because I have to work around the system.
53. I know my supervisor knows that data are sometimes entered incorrectly in datatel.
staff to ensure that data are entered correctly into datatel.
55. If I had my way, CMC would never use datatel.
56. I experience extreme pleasure when entering data in datatel.
57. I'd rather enter data incorrectly than incompletely.

II. Next, in the space below, and only if relevant, please describe a circumstance in which you have entered data into datatel that you knew were incorrect (or later found out were incorrect). In your response, please include the reason you entered the data incorrectly and/or your response to finding out that the data had been entered incorrectly.

III. Thank you for your participation in this research. Before you're finished we'd like to ask you a few questions about yourself. Please indicate your responses to the following items by clicking in the most accurate circle.

59. How long have you been employed at CMC?
1 month or less 1-6 months 6 months to 1 year 1-3 years 3-5 years 5+ years

60. What was your prior experience with data entry before beginning employment at CMC?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| No Experience | | | | | | | | | Highly experienced |

61. On a scale of 1-10, how competent do you feel correctly entering data in the datatel system?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| completely incompetent | | | | | | | | | highly advanced level of competence |

62. (optional) What is your sex?
Male Female

63. (optional) What is your age? (type in your age)
Age:

References

Dekker, S. (2005). *Ten Questions About Human Error: A New View of Human Factors And System Safety.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Dekker, S. (2006). *The Field Guide to Understanding Human Error.* Burlington, VT: Ashgate Publishing Company.

Khan, F., Amyotte, P., & Dimattia, D. (2006). HEPI: A new tool for human error \ probability calculation for offshore operation. *Safety Science, 44 (4),* 313-334.

Koren, E., Goldberg, S., Shinder-Nekhamkin, A. et al. (2007). Study of errors among nursing students. *International Journal of Qualitative Methods, 6 (3),* 71-98.

McNeil, M. & Hoff, D. (2006). Glitches, data errors delay Illinois test-score release. *Education Week, 26(14)* 20-22.

Park, K. & Lee, J. (2008). A new method for estimating human error probabilities: AHP-SLIM. *Reliability Engineering & System Safety, 93 (4),* 578-587.

Shifrin, T. (2007). "Human error behind most serious data loss. Retrieved from www.ecoustics.com/pcw/news/129736 on February 1, 2008.

Tesser, A. (2000). On the confluence of self-esteem maintenance mechanisms. *Personality and Social Psychology Review, 4 (4),* 290-299.

Whittingham, R.B., (2004). The Blame Machine: Why Human Errors Cause Accidents. Boston, MA: Elsevier