XBRL Mandate: Thousands of Filing Errors and So What?

Hui Du
University of Houston – Clear Lake
2700 Bay Area Blvd, Box 42
Houston, TX 77058
281-283-3144
duhui1@uhcl.edu

Miklos A. Vasarhelyi
Rutgers Business School
One Washington Park, Room 946
Newark, NJ 07102-3122
973-353-5002
miklosv@andromeda.rutgers.edu

Xiaochuan Zheng
Bryant University
1150 Douglas Pike
Smithfield, RI 02917
401-232-6811
xzheng@bryant.edu

Acknowledgement: We are grateful for the exchanges with Cliff Binstock, the CEO of XBRL Cloud, who helped us understand the data of XBRL filing errors his company compiled and we used in our study.

XBRL Mandate: Thousands of Filing Errors and So What?

Abstract

Since the mandate by the U.S. Security and Exchange Commission (SEC) to begin the interactive data reporting in June 2009, according to XBRL Cloud, an XBRL product and service provider, more than 4000 filing errors have been identified. We examine the overall changing pattern of the errors to understand whether the vast number of errors may hamper the transition to the interactive data reporting. Using a sample of 4532 filings that contain 4260 errors, we document significant learning curves exhibited by the SEC XBRL filing environment, by the filers, and by the XBRL software vendors. Specifically, we find that the number of errors per filing is significantly decreasing as more quarters pass, when a company files more times, and when a higher version of software is used, suggesting that the SEC, the company filers, and the technology community all learn from their experiences and therefore the future filings are improved. Our findings provide evidence to encourage the regulatory body, the filers, and the XBRL technology supporting community to embrace the new disclosure requirement in financial reporting. The significantly decreased number of errors also helps address the information users' concerns regarding the data quality of XBRL filings.

1. Introduction

In 2009, the SEC issued its mandate requiring that publicly traded companies file their financial statements with the SEC in an interactive data format using the eXtensible Business Reporting Language (XBRL) (SEC 2009a). The mandate is to be implemented over a three year phase-in schedule, beginning in 2009 for the largest companies, and by 2011 for all public companies. Shortly after the completion of the XBRL filings by the first phase-in group of filers, the SEC summarized the XBRL validation errors that will cause the XBRL exhibits not to be filed (SEC 2009b). The SEC staff also identified a number of matters that filers should consider to improve their future filings (SEC 2009c). Similar errors and additional issues in the filing process were discussed by the SEC staff after the second phase-in group completed their filings (SEC 2010). Six quarters after the initial phase-in, XBRL Cloud, an XBRL product and service provider, has compiled more than 4000 filing errors made by more than 1000 filers at its EDGAR Dashboard website.[1]

The objective of this paper is to make a timely examination of the changing pattern of the XBRL filing errors that fail the SEC validations and to draw some inferences on this phenomenon. While prior studies have examined the specific errors made in the XBRL filings (Debreceny et al. 2010; Bartley et al. 2009), we focus on the overall changing pattern of all validation errors in the first six quarters of the mandatory filings that include the first and the second phase-in groups of companies. Since the validation errors cause XBRL filings to be incomplete and inaccurate (SEC 2009b), an analysis of the pattern of the filing errors would help filers understand the error pattern and learn from the errors when they comply with the mandate. The study also provides useful insights to the SEC in evaluating the filing process. It is helpful as

[1]More information about XBRL Cloud is available at: http://www.xbrlcloud.com/

well to the technology community that facilitates the implementation of the new disclosure requirement in that it guides them to better compliance and improved filing practices. Furthermore, the error pattern can help address the information users' concerns regarding the data quality of XBRL filings. The filing data from XBRL Cloud provides us an opportunity to study the errors in a timely fashion after the first and the second groups of phase-in companies have completed their mandatory filings. In the six quarters from the second quarter of 2009 to the third quarter of 2010, we obtain 4532 filings that contain 4260 errors which should not be accepted by the SEC according to its EDGAR Filer ManualVolume II (SEC 2009d).

Based on the learning curve theory in the management science literature (e.g., Argote and Devadas 1991; Ingram 2002; Wiersma 2007), we document significant learning curves measured by the number of errors per filing. The learning curves are exhibited by the SEC XBRL filing environment, by the company filers, and by the XBRL technology vendors. Thus we observe the learning curves in the following three areas. First, the errors are concentrated in the first reporting quarter of each phase-in stage with the first quarter of the first phase-in stage having the highest number of errors per filing. The number of errors per filing lessens in the subsequent quarters. Second, first time filers make the most errors per filing but the errors diminish as these companies file more times. Third, more errors come from the lower versions of software that companies use to file their XBRL documents and the errors decrease when companies switch to higher versions of the same software. Also, the number of errors per filing is higher in quarterly filings than in annual filings because companies typically spend less time on 10-Qs than on 10-Ks. In our regression analysis, after controlling firm characteristics and other confounding factors, we find that the number of errors per filing is significantly negatively associated with the number of quarters passed since the SEC mandate, the number of times filed by a company, and the lower version of software used to file. The decreasing error pattern continues until all four levels of footnote tagging are required one year from a filer's initial filing. We also find that the number of filing errors for a filer is significantly higher in its fifth time filing than in its fourth time filing. The result suggests that the additional footnote tagging requirement starting one year from a filer's initial submission in interactive data triggers a new round of learning curve and cautioning practitioners to pay extra attention when preparing the required four levels of footnote tagging. Moreover, we document positive associations between a company's own creation of XBRL extensions, quarterly filings as opposed to annual filings and the number of errors. Our findings also support the SEC's phase-in schedule that allows the year two and later filers to benefit from the experience of year one filers. In addition, because the number of errors may not be normally distributed across the filings,[2] we use a logistic regression to test the probability of having a filing error or not. The robustness check yields the same results. Collectively, we demonstrate the decreasing pattern of XBRL filing errors among the first two phase-in groups in the first six quarters of the XBRL mandate.

We contribute to the XBRL literature by providing evidence of learning exhibited in the filing process. While thousands of errors in the first two stages of phase-in filings may be of a concern to the parties involved in the XBRL mandate, our findings show the overall error pattern to be diminishing. The findings are consistent with the theory of "learning by doing" in the learning curve literature (e.g., Argote and Devadas 1991; Ingram 2002; Wiersma 2007), suggesting that the SEC, the filers, and the software vendors all learn from their experiences to reduce errors and therefore the future filings are improved.

[2] For example, Debreceny et al. (2010) document that for the computational errors, three quarters of the first round filings in their sample are error free.

Our study is important for the following reasons. First, for the filers who are concerned about the enormous errors in the XBRL filings, we demonstrate the diminished error pattern to alleviate their concerns. This is because companies do learn from experience. Not only do they learn from their own experiences in the XBRL filing process, the so called intraorganizational learning, but also they can benefit from other companies' XBRL experiences, the interorganizational learning. We show evidence to support the SEC's phase-in schedule that allows the year two and later filers to benefit from the experience of year one filers. Even though the interactive data filings are mandatory for all public companies, our findings provide evidence to encourage all companies to embrace this new financial disclosure requirement that is intended to improve the communication and usefulness of financial information. We show that the key factors associated with the filing errors are mainly filing characteristics rather than firm characteristics. This finding should provide the confidence needed for companies to comply with the mandate regardless of their size, industry, operating results, and financial health. Second, the SEC has been monitoring the implementation of the interactive data reporting. It would, "… if necessary, make appropriate adjustments during the phase-in period" (SEC 2009a). If companies are overwhelmed by the vast errors and become frustrated in the filing process, the SEC may need to reconsider the requirement of interactive data reporting or slow down the phase-in schedule. The significantly decreased number of errors over filing quarters provides useful information to the SEC in evaluating the XBRL mandate. Third, our findings are encouraging to XBRL software vendors, practitioners, and filing agents. They actively participate in the filing process and become an equally important part of the learning curves. Our findings suggest that their effort to improve the interactive data reporting by updating their software and using the software more effectively is rewarded by the decreased number of filing errors. Fourth, the evidence of significantly decreased number of errors is helpful in addressing information users' concerns regarding the data quality of XBRL filings. Lastly, we are the first to apply the learning curve theory to the implementation of the XBRL mandate. These interesting learning curves are yet to be explored and investigated further in studying the impact of new technologies on financial reporting.

The remaining paper is as follows. We discuss the background and develop our hypotheses in the next section. The research method is presented in the third section, followed by the discussion of the results in the fourth section. In the last section, we provide the conclusion, limitations, and implications.

2. Background and Hypothesis Development

2.1 Learning curves and errors in XBRL filing

Learning curves have long been documented in the literature of management science, first as rudimentary or anecdotal observations (Wright 1936), and gradually developed into part of the formal theory of organizational learning as empirical or scientific evidence (e.g., Yelle 1979; Argote and Devadas 1991; Wiersma 2007). Knowledge can be learned by units, such as an individual, a group, or a company and the relationships between units in an organization affect learning because knowledge is transferred between units (Argote et al. 2003). Although the management literature has shown a variety of learning curves (e.g., Argote 1993; Wiersma 2007; Morrison 2008), the most classic learning curve captures the mechanism of "learning-by-doing" (Adler and Clark 1991) or experience-based learning. Experience provides learning units with the opportunity to learn through repetition and trial-and-error learning. Individuals and company units learn from their experiences because they absorb and understand knowledge by relating it to what they have done before (Cohen and Levinthal 1990). Typically, the learning curve is steep in the initial stages of learning and becomes less steep in the later stages and may eventually reach a plateau, when additional experience can hardly add more impact on learning (Wiersma 2007). The learning effect of the curves is often measured by company performance such as sales growth, profitability, production costs, or bankruptcy. The empirical studies on those measures suggest that companies benefit from learning-by-doing (e.g., Argote and Devadas 1991; Wiersma 2007). Furthermore, internal firm structure and characteristics can affect the learning curve. Small companies with simple business operations are more likely to learn by doing than large companies with multiple business operations (Ingram and Baum 1997). However, vertically integrated companies with complex business operations learn better from their experience in turbulent environments than less complex companies (Sorenson 2003).

In addition to the intraorganizational learning described above that involves learning within an organization (Argote and Ophir 2002), a company also learns from other companies, the so called interorganizational learning (Ingram 2002). Organizations can benefit from the interorganizational learning because they enjoy the accumulated knowledge without the cost of the accumulated experience (Kraatz 1997). However, the degree of such a benefit depends on the alliance network or competitive relationship among the organizations and how one company applies other companies' experiences (Baum et al. 2000). The three-year phase-in schedule of the XBRL mandate implementation is consistent with the theory of interorganizational learning as the SEC designed the schedule with the intention to "permit companies to plan and implement their data tagging with the benefit of the experience of year one filers" (SEC 2009a).

While learning curves have been largely measured by company performance, the vast errors in the first six quarters of XBRL filing provide a unique setting to study learning curves exhibited by the regulatory body, the filers, and the software vendors. The learning units involved in the XBRL filing process are primarily the groups that are responsible for the filings even though some individuals' effort and learning are inevitable during the process. Specifically, the learning unit at the SEC is the Office of Interactive Disclosure (OID)[3], a filer's learning unit is the company's financial reporting division, and a software vendor's learning unit is its XBRL technology development group. The filing agent used by many companies falls somewhere between a company's financial reporting unit and a software vendor, perhaps more skewed toward a software vendor. While the focus of the learning unit of a software vendor is to develop the XBRL technology, the responsibility of the learning unit of a filing agent (i.e., a project team for XBRL filing) is to use the technology to file interactive data for its corporate clients. These basic learning units exercise the traditional mechanism of learning-by-doing to improve the XBRL filing through repetition and trial-and-error practice.

To measure the SEC's learning curve, we use the number of errors affected by the number of filing quarters passed. After several number of filing quarters have passed, the OID learns from its experiences by reviewing the XBRL data from filers, diagnosing errors, improving taxonomy, and striving to provide guidance and support to filers. Such learning is indicated by the OID's communications with the filers to suggest a number of areas where the filers can make improvement (SEC 2009c; SEC 2010). The filer's learning curve is measured by the number of errors affected by the number of times a filer has filed the interactive data. The filer learns from its filing experience, knowing the mistakes it made before, and avoiding them as the filer tries and repeats the same tasks again. The learning curve of an XBRL software vendor

[3] Established in 2007, OID carries out the SEC's commitment to make financial disclosure accessible and easy to use through the application of interactive data. It helps public companies understand and comply with the XBRL rules for financial statements required by SEC. More information about OID is available at: http://www.sec.gov/spotlight/xbrl.shtml

is measured by the number of errors changed by using a different version of the filing technology the software vendor provides. A software vendor typically replaces a lower version of software with a higher version which incorporates modifications and improvements the software developers have learned from their experience with the lower version.[4] All the three learning curves we measure for the regulatory body, the filers, and the software vendors involve classic experience-based learning.

2.2 Requirement of the phase-in XBRL filings and errors

There has been over one decade of studies and evaluations on the Internet facilitated financial reporting in general and XBRL in particular (Ashbaugh et al. 1999; Debreceny et al. 2002; Ettredge et al. 2002; Debreceny et al. 2005; Kelton and Yang 2008; Debreceny et al. 2010). These studies explored, advocated, and generally supported the idea of using the Internet to facilitate corporate financial disclosure and enhance the communication and usefulness of accounting information. In February 2009, the SEC passed "Interactive Data to Improve Financial Reporting" (SEC 2009a). The mandate requires that companies file financial statements with the SEC and provide such information on their corporate websites in interactive data format using XBRL (SEC 2009a).

For more than one decade beginning in 1993, companies have filed their quarterly, annual, and other reports in HTML (Hyper Text Markup Language) format at the SEC's EDGAR (Electronic Data Gathering and Reporting) database, which automatically collect, validate, index, accept, and forward company filings. The EDGAR database transfers companies' SEC filings from a paper based format to an electronic format and makes them

[4] We do not measure the learning curve of a filing agent, arguing that it is similar to and thus represented by the learning curve of the software vendor.

available to users at the SEC's website. The easily accessible HTML filings on the Internet have greatly improved the distribution of financial information. However, the data in HTML format is not "interactive" in the sense that the data cannot be downloaded into spreadsheets and analyzed with software tools or used within investment models directly without manual re-entering. The new mandate required by the SEC is to transfer its EDGAR database from an old-fashioned document filing system into an interactive data repository, in which company filings are XBRL-tagged and therefore can "interact" with users or be conveniently retrieved in more desirable formats. The interactive data allows financial statement information to be downloaded directly into spreadsheets and analyzed or used in a variety of ways, thus improving the usefulness of accounting information to investors, analysts, and other financial statement users.

The SEC rule mandates a three-year phase-in interactive data reporting for companies' financial statements based on their filing status. It requires that in year one of the phase-in period all large accelerated filers that have a common equity float above $5 billion using U.S. GAAP begin interactive data reporting with their quarterly and annual reports for a fiscal period ending on or after June 15, 2009. In year two of the phase-in period, all other large accelerated filers using U.S. GAAP are required to begin the same interactive data reporting for a fiscal period ending on or after June 15, 2010. All remaining filers, including companies using U.S. GAAP and foreign private issuers using IFRS are required to provide their initial interactive data submission as of June 15, 2011, which is year three of the phase-in period.

Regardless of a company's filing status or the year of its initial phase-in, the first-time filing by a company is different from the rest of its filings in one important way. Within 24 months of the time a company is first required to submit interactive data files, the files are "deemed not filed for purposes of specified liability provisions and protected from liability for failure to comply with the tagging requirements" (SEC 2009a). It gives companies relief by limiting legal liability for the initial filings.

In addition, the new rule requires that footnotes be tagged using four different levels of details, which are (1) each complete footnote, (2) each significant accounting policy within the significant accounting policies footnote, (3) each table within each footnote, (4) within each footnote, each amount. However, in a company's first year of XBRL reporting, only level (1) is required (SEC 2009a) as the four different levels progress and require more tagging details and complexity. Thus, a company only needs to tag each complete footnote as a single block of text at its initial filing and the next three quarterly filings. All four levels of footnote tagging are required starting one year from the company's initial filing, i.e., starting from the company's fifth filing.

2.3 Hypotheses

The errors in the XBRL filings are of great value as they are related to the quality of data and the reliability of accounting information. A prior study has documented the extent and seriousness of filing errors in signage, amounts, labeling, and classification when comparing the XBRL filings by voluntary filers to their 10-K Forms (Bartley et al. 2009). A more recent study using the data from the first quarterly filings by the first group of filers provides evidence about computational errors in the first phase-in quarter of the XBRL mandate (Debreceny et al. 2010). We are interested in examining the overall changing pattern of the errors in the first six quarters of the XBRL filings, which include the relatively straightforward block text footnote tagging and the more challenging requirement of all four levels of detailed footnote tagging.

Based on the classic learning curve theory of learning from experience, the SEC, the filing companies, and the software vendors all learn from the earlier XBRL mandatory filings and use the experience to correct errors and to improve future filings. In addition, companies are shielded by certain safe harbor provisions at their initial interactive data reporting, which may increase the chance of having errors. We predict that the number of filing errors reaches the highest number in the first quarter of the XBRL mandate, at a company's first time filing, and using a lower version of software. The errors gradually decrease when more filing quarters pass, as companies file more times, and if a higher version of software is used. This pattern continues until all four levels of footnote tagging are required one year from a filer's initial filing. This leads to our first hypothesis which consists of three sub-hypotheses:

> **H1a**: *Ceteris paribus,* the number of XBRL filing errors becomes lower as more quarters pass.
>
> **H1b**: *Ceteris paribus*, the number of XBRL filing errors becomes lower as a filer files more times during the first year.
>
> **H1c**: *Ceteris paribus*, the number of XBRL filing errors is lower when using a higher version of XBRL software than when using a lower version of software.

Moreover, all four different levels of detailed footnote tagging are required one year from the initial filings in which only block text footnote tagging is required. These additional tagging requirements demand filers to learn new filing procedures and tasks which trigger a new round of learning curve. We expect a surge in the number of filing errors when a company files the fifth time. After the fifth filing, the number of errors decreases in a repetition of the learning curve. This leads to our second hypothesis:

> **H2**: *Ceteris paribus*, the number of XBRL filing errors for a filer is higher in its fifth time filing than in its fourth time filing.

Learning curves vary considerably depending upon organizational structure, culture, and business stability (e.g., Ingram and Baum 1997). Companies with different firm characteristics may exhibit different learning curves in their XBRL endeavor. Certain characteristics (e.g., simple business operations) may lead to fewer errors in their XBRL filings. Predicting the various error patterns in more detail under a variety of circumstances is beyond the intention and capacity of this study because of the complexity of these learning curves affected and the interference provided by each specific condition (Sorenson 2003). Instead, we control the relevant firm characteristics in our research design.

3. Research design

3.1 Data and sample selection

The data from large scale of XBRL filings has only been available since June of 2009 when the mandatory XBRL filings took effective for the largest filers. Since then, XBRL Cloud Inc. has been collecting and publishing a report (called XBRL Cloud EDGAR Dashboard) on all XBRL filings on a daily basis. We obtained the XBRL filing data for the period from June 2009 to December 2010 from the EDGAR Dashboard website. There is a grace period of an additional 30 days for filers to submit their filings both at their initial filings and the first time they are required to tag footnotes implementing all four levels of details. We have considered this 30-day grace period in selecting our sample.

Table 1 Panel A describes how our final sample was determined. We deleted 25 filings before June 15th, 2009, because they are not mandatory filings. We also excluded 38 duplicate filings. Another 104 filings were eliminated because of missing data in Compustat. Based on these data restrictions, our final sample consisted of 4,532 filings from 1,430 unique companies.

Table 1 Panel B and C present the sample distribution by the number of quarters passed, the number of times filed, and by the creation software. Panel B shows that first time filings account for almost one third (1430/4532) of our sample. In addition, as shown in Panel B, there is a jump in the number of filings in Quarter Five. This is because the second phase-in group

(845 new filers) began to file for the first time in this quarter. Panel C shows that the top five "creation software" account for approximately 95% of all filings.

[Insert Table 1 about here]

3.2 Measure of number of errors

Our measurement of number of errors directly comes from the XBRL Cloud EDGAR Dashboard dataset. In this dataset, an error is defined as "SEC will not (or should not) accept the document according to the EDGAR Filing Manual".

3.3 Test variables

As discussed inSection 2, the three learning curves we examine arethe learning by the regulatory body, the learning by the filers, and the learning by the software vendors. Accordingly, we measure three test variables: the number of quarters passed; the number of times filed and the software version. In addition, considering the additional tagging requirements posed for the fifth time filings, we break down the "number of times filed" variable into two detailed variables: "number of times filed in Phase 1" and "number of times filed in Phase 2".

3.4 Control variables

We classify all the control variables into two categories: firm (filer) characteristics related factors and filing characteristics related factors.

3.4.1. Firm (filer) characteristics related factors

The first determinant under the firm characteristics category is whether or not a filer participated in the voluntary filing program (VFP). Since the VFP participants[5] have already had experience with XBRL as they filed during the program, these companies are more likely to have a higher XBRL mandatory reporting quality.

[5] We obtain VFP participant data from Callaghan and Nehmer (2009).

Firm complexity and business stability also affect learning curves. Small companies with simple business operations generally do better in the learning curve than large companies with multiple business operations (Ingram and Baum 1997). Nevertheless, in unstable business environments, vertically integrated companies with complex business operations learn better from their own experience than companies with simple business operations (Sorenson 2003). Following Doyle et al. (2007), [6] we use FIRM SIZE and FOREIGN TRANSACTIONS to proxy the firm complexity. Since the learning effect depends on various conditions as discussed in the above literature, we do not have directional prediction for the relationships between FIRM SIZE, FOREIGN TRANSACTIONS and the number of errors. To control for business stability and financial health, we use LOSS (whether or not the earnings before extraordinary items is negative) as the measure of financial health and add EXTERME SALES GROWTH and RESTRUCTURING CHAEGE to measure stability (Doyle et al. 2007). We expect there will be a positive association between these three variables and the number of XBRL reporting errors.

3.4.2. Filing characteristics related factors

The second category of the determinants of XBRL reporting errors concerns filing characteristics. The first factor we use is the percentage of extensions (XBRL EXTENSIONS) in a filing. The higher percentage of extensions is expected to be associated with the higher likelihood of errors because companies create their own extensions most likely without the prior tests and validations. The second factor we use is the filing form. A firm usually spends less time on a 10-Q filing than on a 10-K filing. Thus, we expect that a 10-Q related XBRL filing (10-Q FORM) is more likely to have more errors.

---

[6] Interactive data reporting is part of a company's internal control system. We argue that it is appropriate to borrow the definition of these control variables from the internal control literature.

Based on the above discussion, we summarize each of our directional predictions and variable measurements in Table 2.

[Insert Table 2 about here]

3.5 Empirical model

We model the number of errors in an XBRL mandatory filing as a function of the above-mentioned learning curve factors and the firm and filing characteristics using an OLS regression with the following construct:

$$
\begin{aligned}
&\text{\# OF ERRORS=f }(\beta_0 + \beta_1\text{\# of QUARTERS PASSED} + \beta_2\text{PHASE2} \\
&+ \beta_3\text{\# of TIMES PASSED in PHASE1} + \beta_4\text{\# of TIMES PASSED in PHASE2} + \beta_5 \\
&\text{SOFTWARE VERSION} + \sum \beta_n(\text{Control Variable})_n + \\
&\quad {}_j \mathit{INDUSTRY}_j \qquad {}_t\mathit{CREATION} \quad \mathit{SOFTWARE}_t\ ). \qquad (1)
\end{aligned}
$$

All variables are defined in Table 2. In the above equation, we also include industry dummies and creation software dummies to control the clustering effects of industry and creation software.

We test H1a, H1b and H1c by assessing β1, β3, β5 respectively. Based on prior literature, we expect that β1, β3and β5 are significantly less than zero.

To test H2, we need to do a joint test on β2 + β4 – 4*β3. When it is a fifth filing, "phase2" is equal to 1, "# of time passed in phase 1" is equal to 0, and "# of time passed in phase 2" is equal to 1. When it is a fourth filing, "phase2" is equal to 0 "# of time passed in phase 1" is equal to 4, and "# of time passed in phase 2" is equal to 0. Therefore, "β2 + β4 – 4*β3" measures the average difference in # of errors between the fifth time and the fourth time holding all other things equal. Based on prior literature, we expect that "β2 + β4 – 4*β3" is significantly greater than zero.

4. Empirical results

4.1 Univariate test and descriptive statistics

Table 3 Panel A presents descriptive statistics on the key variables. The average number of errors per filings is 0.940 (4260 errors/4532 filings). In addition, the number of errors per filing is skewed to the right with a high variation (standard deviation is 7.202). The average # of quarters passed, # of times filed are 4.321 and 2.631 respectively. Around 88% of filings are 10-Q filings. A very small portion (5%) of our sample is VFP participants. About one third of the sample has foreign transactions.

Table 3 Panel B to D are the breakdowns of the average number of errors by quarters passed, times filed and software versions. Panel B shows that the average # of errors is decreasing until the fifth quarter. We see a similar pattern in Panel C, in which the average # of errors jumped in the fifth time filing. As we expected, this kind of jump in fifth time or quarter is because of the additional tagging requirements posed for the fifth time filings.Table 3 Panel B to D also show the univariate results from t-tests of mean difference across different groups by quarters/times/software versions. In general, most of our univariate results are consistent with prior literature outlined in Section Two. For example, both Panel B and C show that the learning curve was steep in the initial stage of learning, becomes less steep in the later stages, and may eventually reach a plateau.

[Insert Table 3 about here]

Table 4 presents Spearman correlations among the key variables. The first column indicates negative correlations between # of errors and all our three test variables (p-level < 0.01), which is consistent with our prediction. In addition, as illustrated in Table 4, many of our variables are correlated with one another. For example, # of QUARTERS PASSED is significantly correlated with most of the variables. As such, we examine all the determinants further by using multivariate analysis as follows.

[InsertTable 4 about here]

4.2 Multivariate test

Table 5 presents our multivariate test results from the OLS regression equation (1) in Section 3.5 with # of ERRORs as the dependent variable. As shown in Table 5, most of the significant coefficients are in the predicted direction. The coefficient of “NUMBER OF QUARTERS PASSED” is significant (p-value<0.01) and negative, suggesting that the number of XBRL filing errors for a filer becomes lower as more quarters have passed. The coefficient of “NUMBER OF TIMES FILEDED IN PHASE 1” is significant (p-value<0.01) and negative, suggesting that the number of XBRL filing errors for a filer becomes lower as the filer filed more times during the first phase.The coefficient of “SOFTWARE VERSION” is significant (p-value<0.01) and negative, suggesting that the number of XBRL filing errors is lower when using a higher version of XBRL software than when using a lower version of software. Furthermore, the joint test result of Hypothesis Two is consistent with our prediction (p-value<0.10), suggesting that the number of XBRL filing errors for a filer is higher in its fifth time filing than in its fourth time filing.

As for the control variables, the coefficient of XBRL EXTENSIONS is significant (p-value<0.01) and positive, suggesting that the deficiency in XBRL filings is associated with a higher percentage of extensions. The coefficient of 10-Q FORM is significant (p-value<0.05) and positive, suggesting that 10-Q XBRL filings are more likely to have errors when compared to 10-K XBRL filings assuming everything else is equal. The coefficient of VFP PARTICIPANT is positive and significant, which is inconsistent with our prediction, suggesting that VFP is not as effective as the program was intended for.

The correlation chart in Table 4 shows that some variables are highly correlated. To address the concern of multicollinearity, we also examine the Variance Inflation Factor (VIF). None of the variables has VIF (untabulated) greater than 10.

Furthermore, the results in Table 5 provide evidence to support SEC's XBRL phase-in schedule, which allows the year two and later filers to benefit from the experience of year one filers. In the context of our research design, this means that the first time filers in the fifth quarterly filing (i.e., the first filings of the second phase-in group) should have fewer interactive reporting errors than the first time filers in the first quarterly filing (i.e., the first filings of the first phase-in group). Using the results in Table 5, we can infer that on average the first filings of the second phase-in group have 4.716 [7] fewer errors than the first filings of the first phase-in group of filers.

Overall, the results support both H1 and H2, which indicate that the regulatory body, the filing companies, and the software vendors all learn from the XBRL mandatory filings and use the experience to correct errors and improve the future filings.

[Insert Table 5 about here]

4.3 Robustness check

[7] The only difference between the first time filers in the fifth quarterly filing and the first time filers in the first quarterly filing is "number of quarters passed". As shown in Table 5, the coefficient of "number of quarters passed" is -1.179. Therefore, the difference of average error for the first time filers in the fifth quarterly filing and the first time filers in the first quarterly filing is: (-1.179*1) - (-1.179*5) = 4.716.

The descriptive data in Table 3 shows that the distribution of the dependent variable of the above OLS regression model "number of errors" might be skewed with many outliers[8], which suggests that the estimation from the OLS regression might be biased. To address this concern, we use an indicator variable instead of a continuous variable for XBRL filing errors. Accordingly, we model the probability of having an error in XBRL mandatory filing as a function of the above-mentioned learning curve factors and the firm and filing characteristics using a logistic regression with the following constructs:

Prob (ERROR) = f (β_0 + β_1# of QUARTERS PASSED+ β_2PHASE2

+ β_3# of TIMES PASSED in PHASE1+ β_4# of TIMES PASSED in PHASE2+ β_5 SOFTWARE VERSION +$\sum$ a_n (Control Variable)$_n$ +

$$_j INDUSTRY_j \qquad _t CREATION \qquad SOFTWARE_t \text{).} \qquad (2)$$

In the above model, the dependent variable ERROR is an indicator variable that is equal to one if an XBRL filing has an error and zero if a filing has no error. The definition of all other variables is the same as the Equation (1) in Section 3.5.

Table 6 presents our multivariate tests results from the logistic regression equation (2). As shown in Table 6, the coefficients of all three test variables are negative and significant (p-value<0.01), which is consistent with our prediction.

Overall, Table 6 results indicate that our primary results are robust to the different specifications we used in our research design.

[Insert Table 6 about here]

5. Concluding Remarks

[8] We tried different winsorize and the results (untabulated) are not significantly different from the results in table 5.

We examine the overall changing pattern of errors in the first six quarters of the SEC XBRL mandate. Based on learning curve theory, we document significant learning curves measured by the number of errors per filing exhibited by the SEC, the filers, and the software vendors. Consistent with the theory of learning by doing, these learning curves indicate that the regulatory body, the filing companies, the XBRL technology providers, and the filing agents all learn from their experiences and use this knowledge to reduce the errors during the filing process and thus the future filings are improved. In addition, we find that the number of filing errors for a filer is significantly higher in its fifth time filing than in its fourth time filing, suggesting that the additional footnote tagging which is required starting one year from a filer's initial filing triggers a new round of learning curve and thus cautions practitioners to pay extra attention when preparing all four levels of footnote tagging. Our findings also support the SEC's phase-in schedule that allows the year two and later filers to benefit from the experience of year one filers. Overall, our study provides evidence to encourage the SEC, the filers, and the XBRL technology supporting community to embrace the new mandate of interactive data reporting. Also, the evidence of a significantly decreased number of errors provides helpful information to address users' concerns regarding the data quality of XBRL filings.

This study is subject to two limitations. First, we focus on the overall changing pattern of all errors that fail the SEC validations. We do not study the importance of the specific types of errors such as computational errors or XBRL syntax errors, which may be of great interest to researchers and practitioners. Our objective is to examine the learning curves shown by all errors. More detailed studies on the specific types of errors in the XBRL filing process certainly warrant future research. Second, instead of gathering the filing data from the SEC EDGAR database by ourselves, we obtain our data from a publicly available data source. While we rely on the external data in this study, the nature of public data source makes our study easily verifiable and replicable. The data source also provides opportunities for future studies.

In conclusion, despite the large number of errors in the first few quarters of the mandatory interactive data reporting, the pattern of all errors is significantly decreasing. It is believed that the objective of the interactive data reporting to improve the usefulness of financial information is achievable through a collective effort. We provide empirical evidence by examining the learning curves exhibited by the regulatory body, the filing companies, and the technology supporting community.

References

Adler, P. S., K. B. Clark. 1991. Behind the learning curve: A sketch of the learning process. *Management Science.*37 (3), 267–281.

Argote, L., 1993. Group and organizational learning curves – individual, system and environmental components. *British Journal of Social Psychology* 32 (1), 31–51.

Argote, L. and Devadas, R., 1991. Organizational learning curves: a method for investigating intra-plant transfer of knowledge acquired through learning by doing. *Organization Science* 2 (1), 58–70.

Argote, L. and Ophir, R., 2002. Intraorganizational learning. In: *Companion to Organizations*, J. A. C. Baum (Ed.), Blackwell, Oxford, UK, pp. 181–207.

Argote, L., McEvily, B., Reagans, R., 2003. Managing knowledge in organizations: An integrative framework and review of emerging themes. *Management Science* 49 (4), 571–582.

Ashbaugh, H., Johnstone, K.M., Warfield, T.D., 1999. Corporate reporting on the Internet. *Accounting Horizons* 13 (3), 241–257.

Baum, J. A., Calabrese, C. T., Silverman, B. S., 2000. Don't go it alone: Alliance network composition and startups' performance in Canadian biotechnology. *Strategic Management Journal* 21 (3), 267–294.

Bartley, J., Chen, A. Y., Taylor, E. Z., 2009. A comparison of XBRL filings to corporate 10-Ks – Evidence from the voluntary filing program. SSRN working paper.

Callaghan, J., and Nehmer, R. (2009). Financial and Governance Characteristics of Voluntary XBRL Adopters in the US. Working paper, Oakland University.

Cohen, W. M., Levinthal, D., 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35 (1), 128–152.

Debreceny, R.S., Gray, G.L., Rahman, A., 2002. The determinants of internet financial reporting. *Journal of Accounting and Public Policy* 21 (4/5), 371–394.

Debreceny, R.S., Chandra, A., Cheh, J.J., Guithues-Amrhein, D., Hannon, N.J., Hutchison, P.D., Janvrin, D., Jones, R.A., Lamberton, B., Lymer, A., Mascha, M., Nehmer, R., Roohani, S., Srivastava, R.P., Trabelsi, S., Tribunella, T., Trites, G., Vasarhelyi, M.A., 2005. Financial reporting in XBRL on the SEC's EDGAR system: a critique and evaluation. *Journal of Information Systems* 19 (2), 191–210.

Debreceny, R.S., Farewell, S., Piechocki, J., Felden, C., Graning, A., 2010. Does it add up? Early evidence on the data quality of XBRL filings to the SEC. *Journal of Accounting and Public Policy* 29 (3), 296–306.

Doyle, J., Ge, W., and McVay, S. (2007). Determinants of Weaknesses in Internal Control over Financial Reporting. *Journal of Accounting & Economics*, 44, 193-223.

Ettredge, M., Richardson, V.J., Scholz, S., 2002. Dissemination of information for investors at corporate Web sites. *Journal of Accounting and Public Policy* 21 (4/5), 357–369.

Ingram, P. 2002. Interorganizational learning. In: *Companion to Organizations*, J. A. C. Baum (Ed.), Blackwell, Oxford, UK, pp. 644-663.

Ingram, P. and Baum, J. A. C., 1997. Opportunity and constraint: Organizations' learning from the operating and competitive experience of industries. *Strategic Management Journal* 18 (Supplement 1), 75–98.

Kelton, A.S.and Yang, Y.W., 2008. The impact of corporate governance on Internet financial reporting. *Journal of Accounting and Public Policy* 27 (1), 62–87.

Kraatz, M. S., 1997. Learning by association? Interorganizational networks and adaptation to environmental change. *Academy of Management Journal* 41 (8), 621-643.

Morrison, J.B., 2008. Putting the learning curve in context. *Journal of Business Research* 61 (11), 1182–1190.

SEC, 2009a. Interactive Data to Improve Financial Reporting. Securities and Exchange Commission, Washington, DC.

SEC, 2009b. Listing of EDGAR XBRL Validation Errors and warnings. Securities and Exchange Commission, Washington, DC.

SEC, 2009c. Staff Observations from Review of Interactive Data Financial Statements. Securities and Exchange Commission, Washington, DC.

SEC, 2009d. EDGAR Filer Manual (Volume II) EDGAR Filing, 13th ed. Securities and Exchange Commission, Washington, DC.

SEC, 2010. Staff Observations from Review of Interactive Data Financial Statements. Securities and Exchange Commission, Washington, DC.

Sorenson, O. 2003. Interdependence and adaptability: Organizational learning and the long-term effect of integration. *ManagementScience* 49 (4), 446–463.

Wright, T. P. 1936. Factors affecting the costs of airplanes. *J. Aerospace Science* 3, 122–128.

Wiersma, E., 2007. Conditions that shape the learning curve: Factors that increase the ability and opportunity to learn. *Management Science* 53 (12), 1903–1915.

Yelle LE., 1979. The learning curve: historical review and comprehensive survey. *Decision Sciences*10 (2), 302–28.

Table 1
Sample selection

Panel A: Sample selection process

	# of Filings
Initial XBRL mandatory filings from June 2009 to December 2010	4699
Less: Filings before June 15th , 2009	(25)
Less: Duplicate filings	(38)
Less: Missing data in Compustat	(104)
Final sample	4532

Panel B: Sample distribution by number of quarters passed and number of times filed

# of quarters passed	**# of times filed**						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
1	422	0	0	0	0	0	422
2	46	415	1	0	0	0	462
3	6	46	409	3	0	0	464
4	28	5	44	386	3	0	466
5	845	28	6	57	386	7	1329
6	83	830	27	11	57	381	1389
Total	1430	1324	487	457	446	388	4532

Panel C: Sample distribution by creation software

Creation Software	Percentage
Bowne Tagger	23.74
EDGAR Online I-Metrix Xcelerate	22.46
Rivet Software	20.52
Fujitsu XWand	13.59
EDGARizerX	12.95
Clarity FSR	4.37
Others	2.37
Total	100%

Table 2

Variable definitions and expected relation with the number of XBRL reporting errors

Variable	Predicted direction	Definition
NUMBER OF QUARTERS PASSED	-	Measured based on the number of fiscal quarters passed since June 2009
NUMBER OF TIMES FILED	?	Measured based on the number of XBRL filings a company has filed as of the current filing. For example, if a XBRL filing was the second time filed by a company, this variable would be coded as two
PHASE 2	-	An indicator variable that is equal to one if a firm has filed more than four times, and zero otherwise
NUMBER OF TIMES FILED IN PHASE1	-	The number of XBRL filings a firm has done as of the current filing if it is a first phase filing, and zero otherwise
NUMBER OF TIMES FILED IN PHASE 2	?	The subtraction of number 4 from the number of XBRL filings a firm has done as of the current filing if it is a second phase filing, and zero otherwise
SOFTWARE VERSION	-	An indicator variable that is equal to one if the software used was a higher version, and zero otherwise
XBRL EXTENSIONS	+	The percentage of XBRL taxonomy extension in a filing
10-Q FORM	+	An indicator variable that is equal to one if a XBRL filing is a 10-Q filing, and zero if it is a 10-K filing
VFP PARTICIPANT	-	An indicator variable that is equal to one if a firm participated XBRL voluntary filing program (VFP) ,and zero otherwise
FIRM SIZE	?	Log of a firm's total assets at the end of year 2009
FOREIGN TRANSACTIONS	?	An indicator variable equal to one if the firm has a non-zero foreign currency translation [data item #150] in year t, and zero otherwise
EXTREME SALES GROWTH	+	An indicator variable that is equal to one if year-over-year sales growth falls into the top quintile, and zero otherwise
LOSS	+	An indicator variable equal to one if earnings before extraordinary items in year 2009 less than zero, and zero otherwise
RESTRUCTURING CHARGE	+	The restructuring charge in year 2009 scaled by the firm’s year 2009 end market capitalization

Table 3

Descriptive statistics

Panel A: Descriptive statistics of key variables (sample size N=4532).

All variables are described in Table 2.

Variable	Mean	Std Dev	Median	Lower Quartile	Upper Quartile
# of ERRORS	0.940	7.202	0.000	0.000	0.000
NUMBER OF QUARTERS PASSED	4.321	1.663	5.000	3.000	6.000
NUMBER OF TIMES FILED	2.631	1.638	2.000	1.000	4.000
SOFTWARE VERSION	0.604	0.489	1.000	0.000	1.000
XBRL EXTENSIONS	0.124	0.106	0.090	0.050	0.160
10-Q FORM	0.876	0.330	1.000	1.000	1.000
VFP PARTICIPANT	0.049	0.217	0.000	0.000	0.000
FIRM SIZE	8.854	1.630	8.792	7.797	9.866
FOREIGN TRANSACTIONS	0.335	0.472	0.000	0.000	1.000
EXTREME SALES GROWTH	0.201	0.401	0.000	0.000	0.000
LOSS	0.179	0.383	0.000	0.000	0.000
RESTRUCTURING CHARGE	0.004	0.008	0.000	0.000	0.004

Panel B: average # of errors by # of quarters passed

# of Quarters Passed	Average # of Errors	p- value of T test between current quarter and prior quarter
1	8.156	
2	0.606	<0.0001
3	0.103	<0.0001
4	0.099	0.929
5	0.119	0.705
6	0.207	0.092

Panel C: average # of errors by # of times filed

# of Times Filed	Average # of Errors	p- value of T test between current time and prior time
1	2.524	
2	0.286	<0.001
3	0.107	0.038
4	0.105	0.975
5	0.213	0.231
6	0.204	0.935

Panel D: average # of errors by software versions

version	Average # of Errors	p- value of T test between high version and low version
Low	1.599	
High	0.508	<0.001

Table 4

Spearman correlation among key variables

variable	# of ERROR	# of Quar.	#of Times	Version	Extension	Form	VFP	Size	Growth	Foreign	Loss
NUMBER OF QUARTERS PASSED	-0.32***										
NUMBER OF TIMES FILED	-0.21***	0.25***									
SOFTWARE VERSION	-0.12***	0.64***	0.14***								
XBRL EXTENSIONS	0.03**	0.1***	0.39***	0.08***							
10-Q FORM	0.06***	0.18***	-0.19***	0.18***	-0.04**						
VFP PARTICIPANT	0.06***	-0.1***	0.09***	-0.09***	0.07***	-0.03*					
FIRM SIZE	0.14***	-0.35***	0.35***	-0.22***	0.45***	-0.11***	0.14***				
EXTREME SALES GROWTH	-0.03**	0.05***	-0.05***	0.02	0.01	0.03*	-0.08***	-0.09***			
FOREIGN TRANSACTIONS	0.01	-0.02	0.03*	-0.01	-0.07***	-0.02	0.04***	-0.05***	-0.02		
LOSS	0.00	0.04***	-0.04***	0.03*	0.05***	0.01	-0.01	-0.02*	-0.13***	0.04**	
RESTRUCTURING CHARGE	0.02	-0.04***	0.05***	-0.05***	-0.09***	-0.04***	0.07***	0.01	-0.14***	0.24***	0.11***

All variables are defined in Table 2.

(*), (**), (***) indicates significanceat the 0.10, 0.05, and 0.01 levels, respectively.

Table 5
Multivariate analysis: OLS regression

Variables	Pred. Signs	Estimated Coefficients	p-value	
INTERCEPT		5.632	<.001	***
NUMBER OF QUARTERS PASSED	**-**	**-1.179**	**<.001**	*******
PHASE2	**-**	**-2.911**	**<.001**	*******
NUMBER OF TIMES FILED IN PHASE1	**-**	**-0.805**	**<.001**	*******
NUMBER OF TIMES FILED IN PHASE 2	**-**	**0.459**	**0.346**	
SOFTWARE VERSION	**-**	**-1.276**	**<.001**	*******
XBRL EXTENSIONS	+	5.179	<.001	***
10-Q FORM	+	0.826	0.015	**
VFP PARTICIPANT	-	1.303	0.009	***
FIRM SIZE	?	0.100	0.308	
FOREIGN TRANSACTIONS	?	-0.236	0.403	
EXTREME SALES GROWTH	+	0.152	0.529	
LOSS	+	-0.383	0.200	
RESTRUCTURING CHARGE	+	16.456	0.258	
Industry & Software indicator variable		Included		
Number of observations		4532		
Adj R-Sq		0.093		
Test of H2: Fifth Time vs. Fourth time	+	***0.768***	***0.054****	

Dependent variable is number of errors per filing. All other variables are defined in Table 2. (*), (**), (***) indicates significanceat the 0.10, 0.05, and 0.01 levels, respectively.

Table 6
Robustness Check: use dummy variable instead of continuous variable
Logistic regression of the probability of XBRL filing errors

Independent Variables	**Predicted Sign**	**Logit estimate (X^2)**
INTERCEPT		-0.334
		(0.000)
NUMBER OF QUARTERS PASSED	**-**	**-1.133**
		(283.569)*
PHASE 2	**-**	**-1.988**
		(12.924)*
NUMBER OF TIMES FILED IN PHASE 1	**-**	**-0.811**
		(61.283)*
NUMBER OF TIMES FILED IN PHASE 2	**-**	**0.578**
		(1.320)
SOFTWARE VERSION	**-**	**-1.607**
		(61.561)*
XBRL EXTENSIONS	+	5.108
		(33.133)***
10-Q FORM	+	-0.144
		(0.351)
VFP PARTICIPANT	-	0.298
		(1.485)
FIRM SIZE	?	0.026
		(0.218)
EXTREME SALES GROWTH	+	0.031
		(0.028)
FOREIGN TRANSACTIONS	?	0.005
		(0.001)
LOSSES	+	0.004
		(0.000)
RESTRUCTURING CHARGE	+	10.4
		(1.309)
Industry & Software indicator variables		Included
Number of observations		4532
Likelihood-Ratio-Pr>Chi-Sq		<0.001
Pseudo R-Sq		0.254

Dependent variable ERROR is an indicator variable that is equal to one if a XBRL filing has an error and zero if the filing does not have any error. All other variables are defined in Table 2. (*), (**), (***) indicates significance ($P \geq X^2$) at the 0.10, 0.05, and 0.01 levels, respectively.