

Is Using a Form Processing Application Cost Saving? A Real-Life Example

Julien Danhier

GERME (Université libre de Bruxelles), Belgium

Rob Kaelen

GERME (Université libre de Bruxelles), Belgium

Emilie Martin

GERME (Université libre de Bruxelles), Belgium

Alexandra Alarcon-Henriquez

GERME (Université libre de Bruxelles), Belgium

Dirk jacobs

GERME (Université libre de Bruxelles), Belgium

Abstract

A form processing application (such as HP TeleForm) allows transforming scanned forms in an electronic format. The aim of the contribution is to present the whole development of such a solution in a practical project where the team had no previous knowledge on this matter. We discuss the advantages and the limits of the approach, the changes in the way the questionnaire is designed and the data collected. Its costs are also addressed (licenses, hardware, workforce). Moreover, we present the duration of each step of the encoding phase. Finally, the quality of the process is assessed and discussed.

Introduction

In 2014, we conducted a survey among schools in French-speaking Belgium. Data about 10,864 pupils in 104 schools were collected. The 22-page questionnaire consisted of an achievement test in mathematics (61 items) and attitudinal and sociodemographic questions (110 items). The sample size combined with the

Publisher: AAPOR (American Association for Public Opinion Research)

Suggested Citation: Danhier J. 2017. Is Using a Form Processing Application Cost Saving? A Real-Life Example. *Survey Practice*. 10 (2).

ISSN: 2168-0094

questionnaire length would entail an extraordinary workload if we were to encode the data manually. Our estimation of such a workload would not only exceed the duration, but also the budget of our project. In light of this, we considered whether using a data entry tool would be more time and cost-effective.

Based on features of the software and the literature in the medical field (Davidson et al. 1996; Hardin et al. 2005; Nies and Hein 2000; Quan et al. 2003; Wahi et al. 2008), we decided to use HP TeleForm. TeleForm is a form processing application that transforms scanned images into data: a machine-readable form is designed; once the forms are filled in, they are scanned; next, each image is transformed into data through optical recognition technology; finally, the transformation is verified and the data stored in a database. However, this process is not automatic as human intervention is required at each step.

Two main issues guide the choice of a collection procedure: cost effectiveness and reliability. The data capture approach has been chosen based on the assumption that it provides reliable data encoding while it greatly reduces costs. As this chosen alternative to manual encoding is far from being free of charge, such assumptions need to be verified. In order to explore these assumptions, eight classes in four schools (135 students) were randomly selected from our sample – two classes in each of the four socioeconomic strata from our survey. We conducted three types of measures on the subsample of classes:

- All the questionnaires were re-encoded by the same person following both manual encoding (MA) and form processing (FP). Encoding time was measured for each class and costs were estimated.
- Results from both procedures were compared and discrepancies were verified in the questionnaire in order to assess the number of errors produced by each procedure.
- Requests for human intervention were counted in order to identify time consuming questions.

Cost Effectiveness

Table 1 shows the time taken by both “encoding” processes. Clearly, in each class, the time spent for this step is higher for the manual technique (6–11 times higher for the same class). However, this rough comparison is not relevant. Strictly speaking, only a small proportion of FP consists of “encoding” defined as a human using a keyboard to type in or check data. To refine our measure, a cost per questionnaire has been computed for the whole survey and has been added to our current measures.

In FP, forms are scanned; each image is then processed and finally verified by the worker. The software asks for human intervention each time the provided rules do not permit deciding between alternative responses. The previous configuration of rules will define what is ambiguous for the software and consequently, the number of interventions, but also the data quality. With this

Table 1 Time and price.

Class	Size	FP				MA		
		Encoding	Extended encoding	Technical issues	Fixed expenses	Encoding	Technical issues	Fixed expenses
1-1	16	00:10:16	00:22:58 or €5.25	€14.69	€35.43	01:44:01	€24.58	€28.31
1-2	17	00:11:34	00:25:01 or €5.73	€15.77	€37.80	01:34:45	€22.43	€26.39
2-3	15	00:07:43	00:19:39 or €4.48	€13.33	€32.77	01:21:07	€19.21	€22.71
2-4	12	00:07:51	00:17:30 or €3.98	€11.06	€26.61	00:59:40	€14.15	€16.95
3-5	13	00:08:42	00:19:07 or €4.35	€12.02	€28.87	00:57:22	€13.63	€16.66
3-6	20	00:16:30	00:32:15 or €7.42	€19.22	€45.14	01:42:47	€24.36	€29.02
4-7	22	00:11:40	00:28:56 or €6.65	€19.63	€48.14	01:32:47	€22.06	€27.19
4-8	20	00:11:13	00:26:58 or €6.19	€17.99	€43.91	01:25:04	€20.22	€24.89
Total	135	01:25:29	03:12:23 or €44.1	€123.72	€298.68	11:17:33	€160.64	€190.47

extended definition of encoding, the time required to encode one questionnaire is at least doubled. However, the FP remains 1.8–3.9 times faster than the MA.

Compared to MA, using FP entails high extra costs. As some of them are expenses that are not measurable in terms of human time, the price (in euro) required by the encoding of a class given our specific data collection provides an alternative measure. The “extended encoding” serves as baseline measure in euro to make the comparison easier. The time-to-price transformation followed this rule: for a task that can be performed by students, €14 is the hourly wage for 1 hour while this wage reaches €28 for a regular worker.

Next, the resolution of “technical issues” was included. Some issues were expected, others were not, increasing the budget unexpectedly. Among the expected costs, a specific machine-readable form had to be created and tested. The scanner had to be configured to limit the loss of quality. Files had to be gathered in lots to be processed by batches (50 forms by lot, namely 1,100 pages). A database had to be formatted and the export procedure configured (data format, questionnaires saved as pdfs, etc.). In the table, we can see that these technical issues inflate costs (up to 3 times as high) although we can expect a reduction of these cost types as experience increases.

Two examples of unexpected costs are worth noting. The first one concerns the decompression of images for processing. The image format used by the scanner (JPG) was not appropriate and blank pages randomly appeared during the verification step. Consequently, all images were converted to compressed grayscale TIFs (with a limited number of nuances, an optimized contrast and the addition of a slight blur), which solved the problem. The identification of such a problem can be time consuming but provider assistance was helpful. The second problem was deeply time-consuming. Once all the forms were scanned, we ended up having more lines than pupils (duplicated lines and students scattered on multiple lines), due to: non-subsequent pages with the same ID (from disordered or mixed questionnaires); trouble reading some IDs previously written on the forms by the software itself (e.g. 31,220 IDs instead of 11,220, but only for some pages of the questionnaire); alterations made by

or stains added by the scanner and multiple empty lines randomly added in the database. A procedure was developed to group pages with the same ID together to check that no errors remained.

Next, we added what can be called “fixed expenses”. These cover the purchase of the software, a scanner that is able to quickly scan sizable piles of paper and a computer. These costs are high (about €13,500), particularly if they cannot be written off by several research projects.

As regards MA, the time required is larger. However, it suffers from less development or fixed expenses. Only a few hours were required to develop the database and the encoding screens (e.g. formatted spreadsheets), although more complex applications could be developed. Fixed expenses are limited to computers, whose number depends on the amount of time available until the data is needed. In our case, we required that three people work simultaneously on three computers to encode the entire survey in 12 weeks. These limited development costs and fixed expenses only slightly increase the costs of this encoding process.

The comparison of time required by both procedures shows that as long as we do not include fixed expenses, FP is more cost-effective. Moreover, the capitalization of acquired technical know-how in further projects will decrease these costs. However, when the fixed expenses are taken into account, MA becomes more cost-effective. This holds for our specific survey, as a simple computation shows that FP would have become more effective if we had collected data from more than 15,000 pupils. Let us finally note that the FP software failed to read some forms (about 3% of our questionnaires that had to be entered manually).

It is clear that our modeling does not include all the costs and advantages of both techniques. Two non-included costs are worth mentioning. Firstly, the use of FP requires some logistics regarding the printing and dispatching of the questionnaires with their individual ID number (generated to identify each individual form after completion and create lines for each participant in the database), whereas simple copies of the same questionnaire are sufficient when using manual data entry. Secondly, when electronic data capture is used, questionnaires are digitalized, which means paper versions of the questionnaires can be destroyed afterwards, so space (and linked costs) can be saved.

Reliability

Different field types were used (see Figure 1): “multiple choice”, “constrained” and “mixed” fields. In the following paragraphs, we compare the number of errors produced by each procedure and the number of requests for human intervention to identify time consuming questions.

Figure 1 Types of field.

1a. Multiple choice (e.g. gender)

Garçon Fille

1b. Constrained print
(e.g. birthdate)

1 1 / 0 2 / 2 0 0 1

1c. Mixed (e.g. response to a
mathematical problem)

+		2,	50
-	0	0	0
+	1	0	0
	2	0	0
	3	0	0
	4	0	0
	5	0	0
	6	0	0
	7	0	0
	8	0	0
	9	0	0

Multiple choice fields are highly reliable and cost-effective. Constraining the selection of only one choice, requests for intervention are prompted when the software observes more than one filled bullet. If the interviewer describes how to fill in the fields, the software is able to correctly read the replies. Moreover, if the interviewer specifies that a respondent who wants to change his response can simply fill in a second bullet and indicate which bullet is the right one, the right response is easily selected. Finally, by adjusting the sensibility of the software, the stains added by the scanner will not interfere with the reading process. For example, for the gender item, we do not observe any data entry error. Moreover, only one request for intervention was counted, in other words, 0.8% of the valid replies (non-responses have been considered as invalid). Some other multiple choice questions required more interventions (e.g. 5.6% of the valid cases when asking students to identify the higher occurrence on a histogram) but errors were still absent.

However, by allowing the selection of multiple values, the software does no longer prompts for intervention and errors become more difficult to be brought to light. When asking students to identify the two lowest values on a histogram, only two interventions were requested while some errors were not identified for 13 cases (10% of the valid cases). Consequently, this technique can produce many errors, except in the case when the knowledge of the right answer allows routinely identifying data entry errors (e.g. cases including the two correct values among others).

Constrained print fields allow gathering of handwriting information through open questions. However, this format has a low reliability and requires a considerably higher number of interventions. For the date of birth, the software required 103 interventions (83.1% of the valid replies). After interventions, nine data entry errors remained (7.3% of the valid cases). Some numbers are easily mixed up by the software (e.g., 1 and 4). However, note that we observed data entry errors in 1.5% of the valid cases for the manual data entry. In the case of the mother's profession, the software requested interventions in 98% of the valid cases. When the need for human intervention is this high, it does no longer makes sense to compare both processes, as both consist of manual encoding. The only difference is that, with FP, words are pre-encoded but many mistakes generally remain. This pre-encoding could reduce encoding time by allowing post-treatment (removing accents, plural marks or frequent spelling errors) but the number of errors and their diversity make the task difficult.

The mathematical test required a mixed field (Figure 1c) to maximize reliability, as the use of constrained fields has low reliability and requires a lot of interventions and the use of multiple choice to represent numbers is not obvious for respondents. Let us, however, note that this procedure increased print costs and completion time. For the item "Compute $a+1/a$ (if $a=2$)", we did not observe any data entry error. The few inconsistencies between both encoding techniques are not due to the techniques themselves, but to different choices made during both processes to solve discrepancies between the constrained field and the multiple choice one. However, intervention was

requested 43 times (i.e., 41.8% of the valid cases). For this type of field, we observed 20 to 92 intervention requests per item, with a mean of 71.4% intervention requests (41.8% to 79.3% of the valid replies). Notably, FP was quite unable to accurately read the minus sign and, to a lesser extent, the plus sign, even after having improved recognition by making changes regarding the “expected characters” and selecting “machine print characters”. In short, when a constrained print field is required, systematic intervention could be a cautious solution.

In conclusion, the choice of field is of utmost importance as it defines reliability and time consumption. The field has to be chosen with caution to avoid heavy human verification. Although high reliability could be reached by increasing human work, one had to find a balanced solution in the trade-off between time and reliability.

Conclusion

“Is form processing application cost saving?” is not a question to be answered easily. As FP requires an important amount of extra time to prepare encoding (questionnaire design to optimize data recognition, scanning of questionnaires, database configuration) and to recover data (student merging, bug solving), this had to be included in the measure. The best, albeit maybe slightly disappointing, answer we can give is: it depends. It depends on the sample size, the content of your questionnaire and the type of field you use. In our survey, it is far from obvious that FP was cost-saving as our first steps with TeleForm required high fixed and development costs.

However, we are not saying that TeleForm does not do what it is supposed to do. Except for some bugs that we finally resolved, it did the job. Actually, the software is able to do more than we need for a one-shot data collection in the human research field. The question is then: will we use it long enough in order to render it cost-effective? The huge investment needs to be written off. It should be used for more than one research project.

Finally, some limits to our analysis are worth noting. Firstly, the worker that encoded the data for this article was highly efficient. Our measures can be considered as a conservative one. With a less efficient worker, the FP could appear as a more promising option. Secondly, different rules to transform human time to price could provide another result. Thirdly, various questionnaires (regarding length and types of field) can provide various measures. Finally, we were first-time users of TeleForm. A more experienced team would probably spend less time on development and bug resolution although this step will not completely vanish. Therefore, we invite researchers to share their own experience and measures of reliability and cost effectiveness of FP solutions.

Acknowledgements

This work was supported by the European Research Council, under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement 28360, for the EQUOP-project "Equal opportunities in educational systems with high levels of social and ethnic segregation - assessing the impact of school team resources". The authors wish to express their gratitude to Felicia Solis for helpful comments and assistance.

References

- Davidson, L.J., W.J. Ryan, J.M. Rohay and S.M. Sereika. 1996. Technological advances in data entry and verification: is TeleForm for you? *Nursing Research* 45(6): 373–376.
- Hardin, J.M., L.L. Woodby, M.A. Crawford, R.A. Windsor and T.M. Miller. 2005. Data collection in a multisite project: TeleForm™. *Public Health Nursing* 22(4): 366–370.
- Nies, M.A. and L. Hein. 2000. TeleForm: A blessing or burden? *Public Health Nursing* 17(2): 143–145.
- Quan, K.H., A. Vigano and R.L. Fainsinger. 2003. Evaluation of a data collection tool (TELEform®) for palliative care research. *Journal of Palliative Medicine* 6(3): 401–408.
- Wahi, M.M., D.V. Parks, R.C. Skeate and S.B. Goldin. 2008. Reducing errors from the electronic transcription of data collected on paper forms: A research data case study. *Journal of the American Medical Informatics Association* 15(3): 386–389.