

# Towards Higher Quality Data: Impact of Perception of Data Quality on IT Investment - Ghana

<sup>1</sup> Kenneth Kwame Azumah, <sup>2</sup> Henry Osborn Quarshie

<sup>1</sup> MBA, M.Eng. Electrical Engineering & IT, Regent University College of Science & Technology, Accra, Ghana.

<sup>2</sup> MBA (MIS), PG. Cyber Law, Regent University College of Science & Technology, Accra, Ghana

<sup>1</sup> [kazumah@regentghana.net](mailto:kazumah@regentghana.net), <sup>1</sup> [xumah@yahoo.com](mailto:xumah@yahoo.com)

<sup>2</sup> [hquarshie@yahoo.com](mailto:hquarshie@yahoo.com), <sup>2</sup> [sirhoq@gmail.com](mailto:sirhoq@gmail.com)

## ABSTRACT

Quality like beauty is a subjective term that lies in the “eyes of the beholder”. Thus the concept of Data Quality is also mostly intuitive: it depends on the user of the data. The perception of employees about the quality of the data used in their organizations plays a critical role in determining the level of attention given to organizational information. The perception may impact negatively on the level of IT investment needed to build and maintain information systems which are critical in national and organizational governance. Four data quality dimensions namely accuracy, completeness, consistency and timeliness may be measured for any organizational data. A data quality survey could at best reveal the level of usefulness of the data to an employee in his job or to an organization in its operations. In this study, a survey was conducted among employees about the perceived quality of their data based on the four mentioned dimensions of data quality and the data quality initiatives that have been put in place. In the results an average of 66% of the respondents expressed their satisfaction on the data they were collecting, consuming or had in their custody. An even lower percentage of 53% said their organization had adopted some form of data quality management. The results also suggested that the issue of data quality has not been given enough attention by organizations whereas the consumers, custodians and collectors of the information report that there were between 25% to 40% deficiencies in quality of data used in their jobs. A general recommendation was the implementation of data quality initiatives and provision of mechanisms for obtaining feedback from employees on unsatisfactory data items in the workplace.

**Keywords:** Data, quality, perception, quality monitoring, quality initiative, IT investment, e-business.

## 1. INTRODUCTION

Electronic data play a fundamental but vital role in an information age society. Its pervasive nature naturally elicits issues concerning management and “quality”. The concept of data quality is often perceived as one pertaining solely to accuracy however there are other dimensions of quality such as completeness, consistency and timeliness that are relevant across industries where accuracy alone cannot fully describe. Industries such as the stock brokerage and other financial data services may tend to regard as priority, the timeliness of their data alongside accuracy whilst in the engineering industry, consistency with accuracy may be the main concern.

The quality aspect of data is considered at relative importance by organizations and this is reflected by the amount of investment or policies targeted at maintaining or improving it. In business circles, it has been rationalized that data of high quality is a valuable asset, increases customer satisfaction, improves revenues and profits, and offers a strategic competitive advantage [1]. The importance of data quality in both decision and operational processes is also emphasized by [2] who intimated the following as some of the reasons for the growing relevance of data quality in business enterprises: customer matching – where records across several systems are identified as referring to one individual; corporate house-holding – where relationship data between groups of people in a household are leveraged for marketing purposes; organization fusion – where data from the separate arms of a trans-national organization are integrated with consideration for compatibility and

interoperability issues in the resulting data to make them more cohesive and meaningful. The importance of data quality to a business is further underpinned in the intimation by [1] that more chief executives are getting involved in supporting data quality initiatives within organizations.

Despite the afore-mentioned fundamental nature and benefits of high quality data to businesses, there appears to be a dearth of research activity in the area of data quality policy and management among business organizations in developing economies. Ghana, located in the heart of the West African sub-region, has one of the fastest growing mobile-telephony densities in Africa [3] with a fierce level of competition in the mobile phone industry. However the latter is yet to see a significant e-CRM and e-marketing drive that will further accelerate it to higher innovative levels [4]. These businesses however appear to thrive, in spite of the low level of data quality related to marketing and CRM research and consequential cost implications [5]. Also, the area of data quality is given very little attention in developing countries as it is overshadowed by issues which are deemed more pressing within the business.

Providing attention to data quality issues should however serve as a starting point for leveraging the benefits of ICT for the enhancement of business profits. An information system will only be as useful as the quality of the underlying data. Usefulness mostly measured by the Technology Acceptance Model (TAM), is likely to influence data quality improvement initiatives within organizations. As such, perception of the quality of corporate data with regards to the quality dimensions

becomes important in determining the success and benefits of these systems. This study attempts to throw more light on how data quality improvement initiatives can be approached. Little is known about how the perception of data quality impacts on IT investments that are geared towards improving data quality. Data quality awareness and initiatives will be commonplace if the perception of the concept is known.

This study involved a survey conducted to elicit perceptions on various attributes of data quality among 180 consumers, collectors and custodians of data. The results of the survey threw more light on the level of awareness of the need for data quality initiatives to improve the quality of data. About a third of the employees surveyed expressed some level of dissatisfaction about the quality of data being used in their organization. Again about a third of the respondents disagreed that there existed any kind of data quality monitoring and improvement initiative in their organizations. On average it appeared that data quality initiatives if promoted vigorously could impact positively on the overall level of satisfaction experienced by data consumers, collectors and custodians.

## 2. METHODOLOGY

With the aim of getting the perceived relevance accorded data quality issues and the prioritization of quality dimensions within industry, a survey was conducted using printed questionnaires. The population of study covered 250 users of information systems within businesses in Ghana. The geographical location for the conduction of the survey was Accra, a capital city, which clearly had the highest concentration of businesses with information systems. It was also expected that their perceptions would be a fair representation of opinion information on the financial, telecommunication and educational industries, and governmental agencies. The results were analyzed to uncover relationships between data quality dimensions and industry types; perceptions of data collectors, custodians and consumers and the industry type; and perceptions of organizational data quality and relative organizational performance.

The questions were presented in two main formats; short questions with selectable responses on a 1 - 5 Likert scale and closed ended questions with selectable fixed responses. Provision was made for entry of non-available answer to the closed ended questions. The 5-point Likert scale had an undecided category to enable respondents who felt uncomfortable with some of the questions to have an escape route without rendering the questionnaire invalid.

### 2.1 Data Gathering Procedure

Number of questionnaires, administered, returned and problems encountered. A total of 250 questionnaires were issued out of which 180 were returned giving a response rate of about 70%. Out of the returned questionnaires 27 were rejected due to various degrees of inconsistencies in the responses such as

- marking only of extreme positions of the Likert scale
- more than one selection on the Likert scale
- conflicting answers for main and dummy questions

## 3. REVIEW OF LITERATURE

This section examines literature related to data quality, the quality dimensions and the creation and measurement of their metrics.

### 3.1 Data Quality

Data quality is often regarded by researchers as the degree to which data fits its use. Juran [6] refers to high quality data as one that "are fit for their intended uses in operations, decision making and planning. Eppler [7] defines information quality as the fitness for use of information; information that meets the requirements of its authors, users and administrators.

Data quality in itself is a multi-disciplinary area [8] with theories from research on issues such as data quality dimensions, measurement and improvement, and methodologies for data quality policy creation and management [9] [10]. Definition theories attempt to specify the objective and subjective dimensions of data quality through qualitative and quantitative research. Studies [2] suggested that although there were four objective data quality dimensions that cut across all industries, there were subjective dimensions which were related to the type of industry. It is a well-known fact that various kinds of industries have peculiar forms and levels of importance attached to sections of information generated in the particular industry.

This may also imply that for the same industry, a business operating in a developed economy might have different data quality priorities and interests from a business operating in a developing economy [11], with the difference of context arising as a result of the general level of change management and technological competence as applied information technology issues [12]. It follows that motivations and perceptions of whether to invest into data quality initiatives with business resources may thus differ in the two contexts.

The different perceptions of data quality needs has led to the growth of various subjective dimensions of quality metrics over a period of time with many new dimensions being created to fit the ever-changing business needs. According to [1], the three major stakeholders in the evaluation of data quality within an organization, the data collectors, data custodians and the data consumers have different perceptions of the level of data quality within the same organization. Concerning problems associated with data quality, Lee et al. further stated: "Executives are often unaware of data quality problems, or they tend to believe that the information technology department can handle the problems without the allocation of critical time and resources from the top level." [1]

There are documented cases [2] where internal users of data in an organization faced problems mainly of inaccurate and untimely data. Modern research in the area of data quality concentrates on issues of measurement and improvement methodologies [13] [14].

Data collectors, custodians and consumers were surveyed using the information quality assessment instrument [15]. The theory underlying the assessment instrument of user acceptance of information systems or technology is the Technology Acceptance Model (TAM), often times used for accessing the level of acceptance of information systems. The order of acceptance of an information system has been mentioned as based on “perceived ease of use”, “enjoyment” and “usefulness” [16].

### 3.2 Data Quality Dimensions

Data quality dimensions are attributes pertaining to the concept of quality with regards to data or information. It has been realized in this line of research that data quality is contextual in nature, varying in an order of importance across industries. Studies [17] revealed that out of the classifications of data quality mentioned by several authors, six foremost dimensions emerge namely: accuracy, completeness, consistency, timeliness, interpretability and, accessibility. Four of the dimensions have advanced objective measures that can cut across industries and user perspectives and they are accuracy, completeness, consistency and currency.

#### 3.2.1 Accuracy

Accuracy is described in two veins: as syntactic or semantic. Syntactic accuracy determines whether the data item falls within a specified domain using a comparison function. For instance, the first month of the year is “January”; thus in the domain of months of the year, “Janury” may have a lower level of syntactic accuracy. Semantic accuracy defines how close a data item is to its true value in terms of meaning and “truthfulness”. The semantic accuracy goes beyond functional comparison taking the context of the dimensional measurement into consideration [18]. Other studies [19] also support data quality assessment within a contextual framework.

#### 3.2.2 Completeness

Completeness is a dimension that is dependent on the data model from which it is being measured. It can be measured based on a population check for completeness in a column of a table containing data. It is normally measured as a ratio of the acceptable data items to the total number of data items in the source [18].

#### 3.2.3 Timeliness

Time-related dimensions describe promptness, freshness and frequency of updates of data. The dimensions include currency, timeliness and volatility [18].

#### 3.2.4 Consistency

Consistency deals with the violation of semantic rules defined over data items. It determines if a unit of data is specified the same way throughout an information system. This goes to buttress the point that to fully understand the problem of data quality, its dimensions need to be studied [18].

## 4. RESULTS AND DISCUSSION

The results obtained are grouped into two charts. Figure 1 shows the responses on data quality dimension related questions and figure 2 shows the responses to the perception of the data quality initiatives embarked upon by organizations. In reporting the perception of employees on the quality of their organizations’ data, it was also vital to note the level acceptance of the information systems that generate information based on the data.

### 4.1 The Industry, Ease of Use of the Information System and the Usefulness of Data

Questions that addressed this concern elicited the usefulness of the information system to the employees’ work and also obtained the ease of use of the system. Thus in order to determine the level of usefulness of the data, the information system itself must be easy to use.

About 82% of the users find their systems easy to use and about 79% intended to continue using the system for another year giving a reasonably high percentage of users who were comfortable with their organizations’ information systems.

Out of the responses obtained 66% were primarily data consumers, 13% were collectors and 21% were custodians. The respondents represented various industries with the largest percentages being the IT/Research (26%), Education (17%) and Financial (16%). The kind of data being consumed, collected and kept was mainly customer/client information (37%), financial information (20%) and human resource data (14%).

### 4.2 Perception on Data Quality Dimensions

About 75% of the respondents agreed that their data included the necessary values for completeness and 82% agreed that the data was in sufficient volume to support their work. Seventy-five percent said information in the system was accurate. Eighty-seven percent reported that the information was accessible and 83% agreed it was easily retrievable. That the data was presented in a consistent manner was agreed upon by 64% of the respondents and 68% also said it was believable. Ninety-two percent said the information was useful for their work. The timeliness of the information was agreed to by 80% but 72% then reported that it was sufficiently up-to-date for their needs.

### 4.3 Data Quality Initiative and Organizational Tools for Monitoring

With the adoption of some form of data quality management, 71% agreed their organizations had made an effort at some data quality initiative. However a reduced

<http://www.cisjournal.org>

63% agreed that their organizations had the tools to identify the deficiencies in its information. Fifty-nine percent said that the organization had tools to ensure the consistency of its information and a similar 57% and 59% agreed that their organization had the tools to ensure the completeness and correctness respectively of its information.

Seventy percent of the valid responses said that their organization had a unit responsible for information quality. However only 55% and 60% said their organization had tools for ensuring the completeness and accuracy respectively of its information.

Also a significantly low percentage (45%) said their organization was developing a dictionary for standardization of its data across different computers, branches and divisions.

Sixty-three percent of employees reported that they were empowered to provide feedback on the information they used in their work.

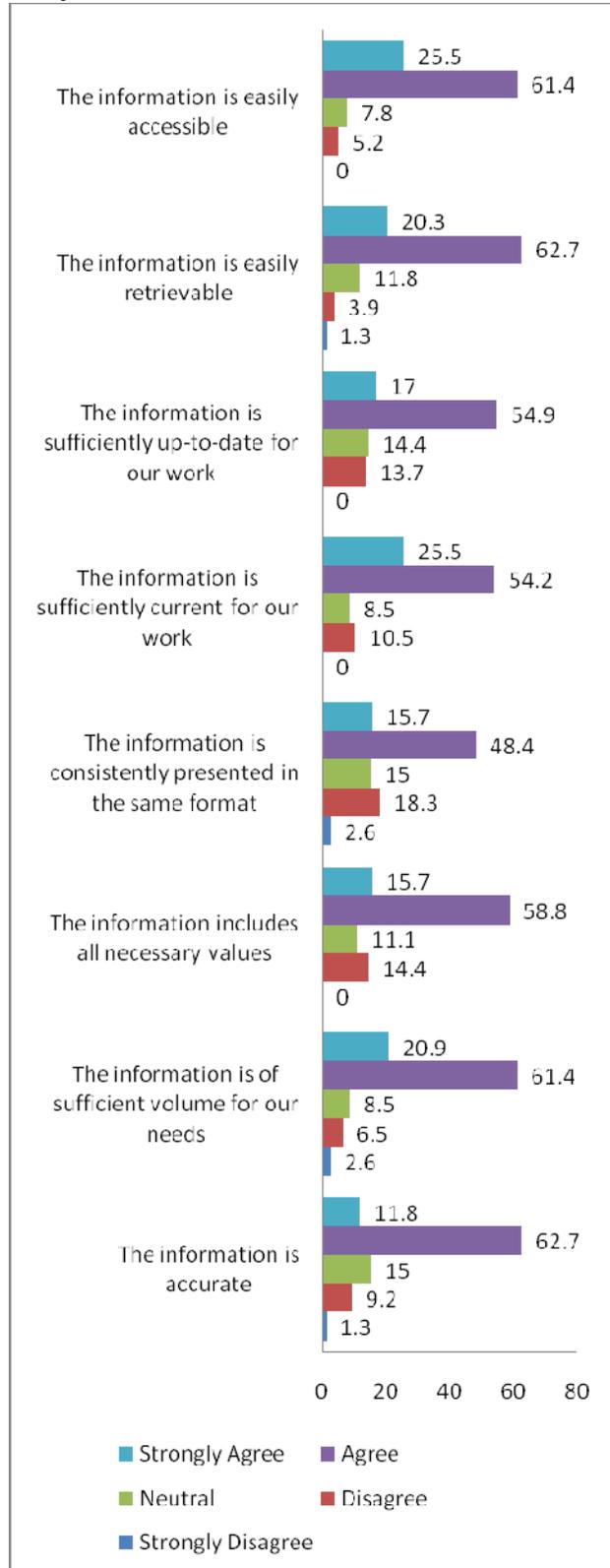


Fig 1: Perception on Data Quality Dimensions

http://www.cisjournal.org



Fig 2: Perception on Data Quality Initiatives

## 5. CONCLUSIONS

An average of about 66% of the respondents expressed satisfaction on the data they were collecting, consuming or had in their custody. An even lower percentage of 53% said their organization had adopted some form of data quality management. The percentages suggest that the issue of data quality has not been given enough attention by organizations whereas the consumers, custodians and collectors of the information report that

there are between 25% to 40% deficiencies in quality of data used in their jobs.

## 6. RECOMMENDATIONS

Data quality initiatives involve activities that aim at improving the quality of organizational data by providing a feedback mechanism for employees to report data quality deficiencies. This is effectively done by creating an organizational unit that will be responsible for collecting all the feedback and updating or acting upon the low quality data. The unit must be adequately resourced to function effectively and some form of education carried out within the entire organization to facilitate the free reporting of any unsatisfactory quality of data.

The information systems should include a feedback functionality that will facilitate the "tagging" of piece of information that is deficient in any predefined quality criteria or at least in any of the four objective quality attributes. A usual first step to identifying data quality deficiencies involves running a profiling tool. The profiling software must be run regularly to provide an idea of whether the initiatives are impacting on the overall quality of data and also to identify new deficient areas of quality that would need attention. Data consumers are then encouraged to provide feedback or are empowered to make corrections to the data.

A data administrator could also be engaged to coordinate the feedback process. In large multinational organizations, an electronic means of submitting deficiencies has to be adopted. More importantly, the CEOs must be supportive or even initiators of data quality management within their organizations before any form of sustainable improvement could be attained.

## REFERENCES

- [1] Y. Lee, L. Pipino, J. Funk and R. Wang, *Journey to Data Quality*, London: MIT Press, 2006.
- [2] C. Batini and M. Scannapieco, *Data Quality – concept methodologies and techniques*, New York: Springer, 2006.
- [3] Ministry of Communication, Ghana, "Boost in t Telecom Industry," 2006. [Online]. Availab: <http://www.moc.gov.gh/print.php?sid=166>. [Access 24 September 2008].
- [4] Derfler, J. F; Editors of PC Magazine , *e-Busine Essentials: Successful e-Business Practices*, New Yor: Que Publishing, 2000.
- [5] T. Redman, "The Impact of Poor Data Quality on t Typical Enterprise," *Communications of the ACM*, v 41, pp. 79-82, 1998.
- [6] J. M. Juran, *Juran's Quality Handbook*, New Yor: McGraw-Hill, 2000.

<http://www.cisjournal.org>

- [7] M. J. Eppler, *Managing information quality : Increasing the value of information in knowledge-intensive products and processes*, Berlin: Springer, 2003.
- [8] W. Chung, C. Fisher and R. Wang, "What Skills Matter in Data Quality?," 2003. [Online]. Available <http://web.mit.edu/tdqm/www/tdqmpub/WSMDQ-ICIQNov02.pdf>. [Accessed 2 May 2008].
- [9] R. Wang, V. Storey and C. Firth, "A Framework for Analysis of Data Quality Research," *IEEE Transactions on Knowledge and Data Engineering*, vol. 7, pp. 62-640, 1995.
- [10] R. Wang, H. Kon and S. Madnick, "Data Quality Requirements Analysis and Modeling," in *Proceedings Ninth International Conference on Data Engineering* Vienna, 1993.
- [11] E. Shih, K. Kraemer and J. Dedrick, "IT Diffusion in Developing Countries," *Communications of the ACM* vol. 51, pp. 43-48, 2008.
- [12] A. Salman, "Elusive challenges of e-commerce management in developing countries," *Business Process Management Journal*, vol. 10, no. 2, pp. 140-157, 2004.
- [13] I. Agosta, "Trends in Data Quality," *DMReview*, vol. 15, pp. 34-36, 2005.
- [14] V. Raeburn, "Measuring Customer Data Quality," *DMReview*, vol. 17, pp. 34-35, 2007.
- [15] CRG (Cambridge Research Group), "Information Quality Assessment (IQA) Software Tool," Cambridge Research Group, Inc, Cambridge, Mass., 1997.
- [16] H. Van der Heijden, "User Acceptance of Hedonic Information Systems," *MIS Quarterly*, vol. 28, no. 4, pp. 695-704, 2004.
- [17] C. Cappiello, C. Francalanci and B. Pernici, "Data quality assessment from the user's perspective," in *IQI Paris*, 2004.
- [18] C. Batini and M. Scannapieco, "Data Quality Dimensions," in *Data Quality – concepts, methodologies and techniques*, New York, Springer, 2006, pp. 21-31.
- [19] A. Even and G. Shankaranarayanan, "Utility-Driven Assessment of Data Quality," *The DATA BASE for Advances in Information Systems*, vol. 38, pp. 75-90, 2007.

## APPENDIX

### QUESTIONNAIRE

#### Section 1: Information System Utility and Acceptance

For each statement, indicate the extent to which you agree about the information SYSTEM you use in your organization. These include spreadsheets, accounting, banking, HR, CRM software, etc. Please circle/tick the number that best suits you.

Strongly Disagree (SD) = 1; Disagree (D) = 2; Undecided (U) = 3; Agree (A) = 4; Strongly Agree (SA) = 5

	Question Item	SD	D	U	A	SA
1	A specific person (or group) is available for assistance with system difficulties.	1	2	3	4	5
2	I intend to use the system for the next 12 months	1	2	3	4	5
3	Overall, I find the system easy to use.	1	2	3	4	5

**Section 2: Dimensional Information Quality Assessment**

Please tick one option that best suits you.

**1. The primary type of the organization's information that I use is mainly**

- |   |   |
|---|---|
| <input type="checkbox"/> Financial or Accounting Data     | <input type="checkbox"/> Production or Manufacturing Data |
| <input type="checkbox"/> Customer, Client or Patient Data | <input type="checkbox"/> Marketing or Sales Data          |
| <input type="checkbox"/> Human Resources Data             | <input type="checkbox"/> Other (Specify) _____            |

**2. How will you primarily describe your role?**

- |   |   |
|---|---|
| <input type="checkbox"/> Collect this information                   | <input type="checkbox"/> Use this information in my tasks |
| <input type="checkbox"/> Work as an information system professional | <input type="checkbox"/> Other (Please specify) _____     |

**3. How will you primarily describe your industry?**

- |  |                                      |                                    |   |
|--|--------------------------------------|------------------------------------|---|
| <input type="checkbox"/> Agricultural          | <input type="checkbox"/> Educational | <input type="checkbox"/> Financial | <input type="checkbox"/> Information Technology |
| <input type="checkbox"/> Manufacturing         | <input type="checkbox"/> Mining      | <input type="checkbox"/> Health    | <input type="checkbox"/> Telecommunication      |
| <input type="checkbox"/> Other (Specify) _____ |                                      |                                    |   |

For each statement, indicate the extent to which your primary type of information characterizes your organization and its quality activities. Please circle/tick the number that best suits you.

		SD	D	U	A	SA
1	The information is consistently presented in the same format	1	2	3	4	5
2	The information includes all necessary values	1	2	3	4	5
3	The information is easily retrievable	1	2	3	4	5
4	The information is accurate	1	2	3	4	5
5	The information is believable	1	2	3	4	5
6	The information is of sufficient volume for our needs	1	2	3	4	5
7	The information is protected with adequate security	1	2	3	4	5
8	The information is useful to our work	1	2	3	4	5
9	The information is easily accessible	1	2	3	4	5
10	The information is sufficiently current for our work	1	2	3	4	5
11	The information is incorrect	1	2	3	4	5
12	The amount of information does not match our needs	1	2	3	4	5
13	The information is not sufficiently timely	1	2	3	4	5
14	The information is incomplete	1	2	3	4	5
15	The information is sufficiently complete for our needs	1	2	3	4	5
16	The information is sufficiently up-to-date for our work	1	2	3	4	5
17	The information is reliable	1	2	3	4	5

**Data Quality Initiatives or Programmes**

		SD	D	U	A	SA
1	My organization has adopted some form of data quality management	1	2	3	4	5
2	My organization has the tools that identify deficiencies with its information	1	2	3	4	5
3	My organization has tools to ensure the consistency of its information	1	2	3	4	5
4	In this company, there are people whose primary job is to ensure the quality of information	1	2	3	4	5
5	My organization has a specific person or group responsible for	1	2	3	4	5

<http://www.cisjournal.org>

	information quality					
6	My organization has tools to ensure the completeness of information	1	2	3	4	5
7	In this company, employees are able to take actions to improve the quality of information	1	2	3	4	5
8	My organization is developing a data dictionary to standardize data definitions across different computers, branches or divisions	1	2	3	4	5
9	My organization has tools to ensure the correctness of its information	1	2	3	4	5
10	This company has recently moved this information to a different hardware or software system	1	2	3	4	5
11	In this company, ensuring the quality of the information is the responsibility of those who use the information	1	2	3	4	5
12	Poor data quality is an IT problem	1	2	3	4	5
13	Government policies affect our data quality programmes or initiatives.	1	2	3	4	5