

**Scorecard Spot Check Evaluation
BENAZIR INCOME SUPPORT PROGRAM (BISP)**

**DATA ENTRY SPOT CHECK
PHASE SEVEN REPORT
June, 2013**



Table of Contents

Introduction.....	1
Background.....	1
Methodology.....	1
Sample Size.....	2
Analysis tools.....	2
Implementation	4
Work Schedule.....	4
Logistics.....	4
Data Base Development and Data Entry	4
Monitoring and Supervision of Data Entry.....	5
Hiring and Training of Staff	6
Hiring of Staff.....	6
Training.....	6
Analysis and Findings.....	7
NADRA’s Data Entry Methodology	7
Discrepant Households	7
Frequency of Errors	8
Calculation of PMT Score	9
PMT Score Discrepancy	10
Discrepancy in PMT Scores of Non-Discrepant Households.....	12
Findings.....	13
Conclusion	14
Annex 1: DEO wise discrepancy in questions.....	15

List of Tables

Table 1: Data Entry Spot Check Phase 5 Sample	2
Table 2: Data Entry Spot Check Phase 5 Sample-DEO Wise	2
Table 3: Number of Discrepant Households.....	8
Table 4: Number of Non Discrepant Households.....	8
Table 5: Number of Discrepant Questions.....	8
Table 6: Number of Errors per Question	9
Table 7: DEO Wise PMT Score Calculation	10
Table 8: PMT score Discrepancy.....	11
Table 9: DEO wise Difference in Score Range	12
Table 10: Non-discrepant households with Discrepancy in PMT scores	12
Table 11: Difference in data entry for the DEOs.....	13

List of Figures

Figure 1 Discrepant Households.....	7
Figure 2: Score Discrepancy	10
Figure 3: Score Difference Range.....	11

Introduction

Background

The Benazir Income Support Program (BISP) was launched in 2008 by the Government of Pakistan as the country's primary social safety net. The idea behind this initiation is to counter the effects of rising food and energy prices on poorer households. The BISP intends to give a cash grant of PKR 1,000 per month to deserving poor families. Since an additional purpose of the program is to empower women, therefore only the adult (above 18) female(s) in a household are eligible to receive the cash grant. Eligibility is determined through the calculation of Proxy Mean Test (PMT) score. Those falling below a predetermined cut off point are determined as eligible to receive benefits through the program.

For this purpose households are surveyed by Partner Organizations (POs). The POs hand over all collected information (T1 forms) to NADRA Headquarters, Islamabad. These are scanned and sent for data entry across the country to the contracted Data Entry Organizations (DEOs). The forms are entered in a MIS developed specifically for this program. This MIS allows for entries such as names, CNIC, address, etc to be verified with NADRA's database. The software calculates the PMT scores of households and houses below the agreed PMT score are identified.

Methodology

IDS has been contracted by BISP to assess the accuracy of data entry conducted by NADRA. This study evaluates the performance of the DEOs contracted by NADRA for data entry. For this purpose a sample (batches) of scorecards selected from those completed by various Partner Organizations (POs) who have been contracted to collect the scorecard information by BISP, are entered for each of NADRA's Data Entry Organization (DEO), by IDS into a MIS system developed specifically for this purpose. This data is then compared with the DEO entered data, to establish accuracy of data entry. The purpose of this component of the spot check evaluation is to determine the performance of the DEO and the MIS. Batches which fall within a pre-defined error margin are deemed to be accepted. Those that do not, will be re-entered by the DEO.

The whole activity is to be divided over eight different phases out of which six phases have been completed. Phase Seven and Eight for the Data Entry spot check began when IDS was provided scanned copies of the 5,500 forms by BISP on April 30, 2013.

The specific objectives of the data entry spot check are as follows:

- Test the accuracy of data entry: determine the frequency of incorrect entries
- Evaluate the performance of the DEOs¹
- Check to see if there are systematic errors e.g. if the frequency of error is higher for particular questions or if frequency of errors are higher in particular offices of the DEOs
- Identify the reasons behind discrepancy in data entry

¹ Performance of DEO refers to how accurately the data entry stations enter data so that a comparison can be made between them.

Sample Size

The total sample size for the Data Entry Spot Check is 22,000 households. These were to be divided among 8 phases, i.e. approximately 2,750 households in each phase. The sample size for the Data Entry Spot Check phase 7 and 8, i.e. 5,500 households, was drawn from the 17,577 matched households of the Targeting Survey Spot Check Phase 3. The sample for Phase 7 of the Data Entry Spot Check has been drawn from matched households from the following districts, as shown in Table 1.

Table 1: Data Entry Spot Check Phase 7 Sample

District	Number of Households
Bahawalnagar	694
Sargodha	729
Kohat	157
Mandi Bahuddin	260
Khyber Agency	102
Ghotki	483
Khairpur	357
Total	2782

NADRA provided IDS with the sample divided over 9 DEOs for comparison of performance across the different DEOs. The sample size of each DEO was selected in proportion to the number of questionnaire each had entered from the matched Targeting Survey Spot Check Phase 3 households.

Table 2: Data Entry Spot Check Phase 7 Sample-DEO Wise

DEO	Number of Households
Adv. E-Tech	135
Deloitte	305
DPS	490
IA	318
MYASCO 360	235
NCBMS	81
NIFT	662
Systems	503
Others*	53
Total	2782

*Others represent several DEOs with a very small sample size. These include HQ, PHQ Karachi, RHQ Multan, RHQ Sukkur and others.

Analysis tools

Data entered by IDS is matched and compared with data entered by the DEOs. Indicators have been formulated to measure the extent of discrepancies/incorrect entries and identify their source. Analysis is conducted using indicators that look for systematic errors and variability in accuracy across offices (DEOs). As such, the following indicators are used:

- **Question Indicator:** This indicator measures the percentage of incorrect entries to determine if particular questions have heightened inaccuracy.
- **DEO Indicator:** This indicator measures the percentage of incorrect entries by each DEO in order to identify DEOs with higher errors.
- **PMT Score Indicator:** The percentage of households with difference in score calculated by IDS and NADRA/DEO.

This is the Data Entry Spot Check Phase 7 Report which is provided as part of the overall deliverables.

Implementation

Work Schedule

The start-date for each phase of the Data Entry spot check is dependent on when the sample questionnaires are made available. The start date proposed is when the data entry by the DEOs and validation at NADRA is expected to be completed for any cluster. Scanned copies of forms of the selected beneficiaries for this phase were received on April 30, 2013.

Logistics

Project Coordinator (Operations) is the overall in charge of the whole of Data Entry Spot Check activity. All communication with BISP Headquarters and NADRA including transfer of data, reports at required interval and other deliverables take place through the Project Coordinator (Operations). The IDS head office supervises the overall activity and performance of the team members. The MIS Manager is responsible for managing all tasks that involve data at various stages. His major responsibilities include: receiving data from the BISP office, development of software for data entry and processing, testing of software, supervising the key punch operators (KPOs) and data editors in data entry and cleaning process, processing data to ensure accuracy and readability to carry out further analysis including the indicators defined in the preceding section.

Key Punch Operators (KPOs) are responsible for data entry into the software specially designed for this activity. KPOs work in close coordination with data editors and MIS Manager. The KPOs hired for Phase 5 were the same as in the previous phases. These KPOs had already gone through the three days training workshop and had been tested by holding a mock data entry exercise using the developed software in order to qualify for the real task. Since the KPOs had already attended the training sessions, they went through a one day refresher for this phase. Software data editors are responsible for reviewing and cleaning data entered by the KPOs and providing them feedback on their performance in order to rule out human error at data entry stage at IDS. Data analysts work in close coordination with the MIS Manager and department in generating the indicators defined and report writing.

Data Base Development and Data Entry

A database has been created at IDS using SQL Server 2000. Data entry is carried out on the basis of double entry and checked carefully to ensure near perfect accuracy providing a strong base against which to compare the DEOs' data entry. When a form is entered once by a KPO, a unique key is generated, and a colored tag is placed on the form which has information about the name of the KPO, identification code of the KPO who entered the form into the software, source of data (office) which in this case is NADRA, number of times the form has been entered into the software i.e., first or second entry, unique key generated by the software on completion of each form, survey phase, quarter number and date of data entry. This is to ensure that each form is entered twice and the unique key ensures traceability of the form in case errors during the data entry need to be corrected. The forms entered twice, as indicated by the information completed on the tag are passed on to the MIS department.

Monitoring and Supervision of Data Entry

Once the data had been entered into the software, editors in the MIS department review the data entered of each part of the T1 form in order to clean data of any data entry errors. For further verification, both data sets are transferred to SPSS (at random intervals) in order to allow for a comparison of the software. This allows any bugs in the software to be detected. Once the data is verified, it is made available for analysis. The MIS manager then works in close coordination with the data analysts to get the required outputs for the reports.

Hiring and Training of Staff

Hiring of Staff

All staff hired for the Data Entry Spot Check had at least a bachelor degree; preference was given to staff from IDS's existing roster. A total of 25 Key Punch Operators (KPOs) worked on a full time basis for the period under report. Additionally, IDS hired Quality Control Officers (software) who were responsible for cleaning the data entered by the KPOs and providing feedback on performance in order to minimize human error.

Training

As already stated IDS organized a one day refresher session for the KPOs and Quality Control Officers (QCOs) at the IDS head office on May 12, 2013. The KPOs hired for this phase were the same as the previous phase and were familiar with the questionnaire and the software. The purpose of the one day refresher was to review the understanding of the questionnaire, data entry software and different quality/security protocols for data entry.

Analysis and Findings

NADRA's Data Entry Methodology

NADRA calculates the age of household members according to the rule:

“If Date of birth is given then age is calculated with following formula $DOB - Current Fiscal Year (2011-07-01)$, otherwise given age is considered”

Room Ratio is a ratio of the number of rooms to the number of household members. As per instructions issued by The World Bank, the total number of household members was to be calculated from the household roster. However, as confirmed, NADRA considers the number of household members as entered for question 24(back side of the questionnaire) when calculating the room ratio².

The analysis in this report is based on NADRA's data entry methodology.

Discrepant Households

A discrepancy is identified when there is a difference between data entered for a question by NADRA/DEO and data entered for the same question by IDS. A discrepant household is a household for which there is a discrepancy in at least one question. As the figure 1 shows, overall there were 5.3 percent discrepant households.

Figure 1 Discrepant Households

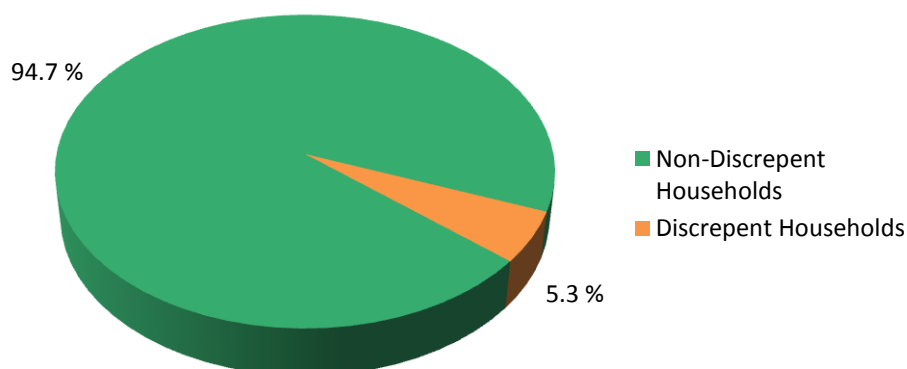


Table 3 shows the DEO wise percentage of data entry errors. For all DEOs except Systems the percentage of this error was more than 3 percent. It was higher for NCBMS and Adv.E-Tech, i.e. 13.6 percent and 12.6 percent of their respective samples with errors in data entry.

² The number of household members from the household roster and question 24 should be same. However, there were cases where these did not match, identifying enumeration error.

Table 3: Number of Discrepant Households

DEO	Number of households surveyed	Number of Discrepant Households	Discrepant Households (As a Percentage of Total Households Interviewed)
Adv. E-Tech	135	17	12.6%
Deloitte	305	14	4.6%
DPS	490	24	4.9%
IA	318	15	4.7%
MYASCO 360	235	21	8.9%
NCBMS	81	11	13.6%
NIFT	662	29	4.4%
Systems	503	14	2.8%
Others	53	2	3.8%
Total	2782	147	5.3%

Table 4 shows the number of non-discrepant households. Data entry carried out by the DEOs was accurate for 94.7 percent of the selected households.

Table 4: Number of Non Discrepant Households

DEO	Number of Households Interviewed	Number of Non-discrepant Households	Non-Discrepant Households (As a Percentage of Total Households Interviewed)
Adv. E-Tech	135	118	87.4%
Deloitte	305	291	95.4%
DPS	490	466	95.1%
IA	318	303	95.3%
MYASCO 360	235	214	91.1%
NCBMS	81	70	86.4%
NIFT	662	633	95.6%
Systems	503	489	97.2%
Others	53	51	96.2%
Total	2782	2635	94.7%

Frequency of Errors

The frequency of errors is measured by the number of questions with erroneous data entry. Table 5 summarizes the number of households for different number of errors. Of the total discrepant households 89.1 percent (131 of 147) had errors in the data entry of one question. The maximum number of data entry errors for a household was in four questions. There were four such cases.

Table 5: Number of Discrepant Questions

DEO	Number of Questions with Discrepant Answers in Each Household			
	1	2	3	4
Adv. E-Tech	14	3	0	0
Deloitte	13	1	0	0
DPS	23	1	0	0
IA	13	2	0	0
MYASCO 360	20	0	1	0
NCBMS	10	1	0	0
NIFT	24	2	0	3
Systems	12	1	0	1
Others	2	0	0	0
Overall	131	11	1	4

Table 6 shows the number of errors for each question. Most of the differences identified were in the data entry of the number of dependents and children’s education. The number of dependents and children’s education did not match for 61.9 percent and 12.9 percent, respectively, of the 147 discrepant households. Both of these are not directly taken from the questionnaire but depend on the age calculated of the household members listed in the roster³. IDS followed the methodology as shared by NADRA for the calculation of age. Hence, the discrepancy in these variables can be attributed to errors in data entry.

Table 6: Number of Errors per Question

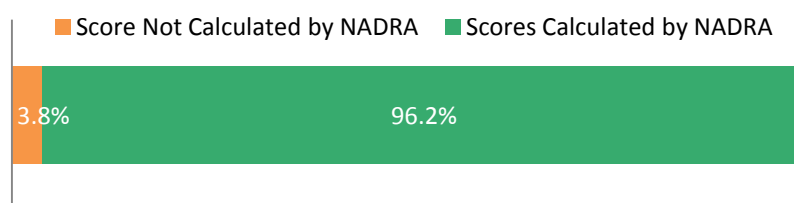
Question	Number of Households with Discrepancy
Discrepancy in Number of Dependents	91
Discrepancy in Children Education	19
Discrepancy in Goat Ownership	6
Discrepancy in Freezer Ownership	3
Discrepancy in Microwave Oven Ownership	2
Discrepancy in Motorcycle Ownership	2
Discrepancy in Car Ownership	2
Discrepancy in Household Head Education	1
Discrepancy in Washing Machine Ownership	1
Discrepancy in Cooking Stove Ownership	1
Discrepancy in Cooking Range Ownership	1
Discrepancy in Geyser Ownership	1
Discrepancy in TV Ownership	1
Discrepancy in Bull Ownership	1
Discrepancy in Buffalo Ownership	1
Discrepancy in Sheep Ownership	1
Discrepancy in Land Unit	1

When analyzed across DEOs, number of dependents and children’s education remain questions with the most discrepancies. (See *Annex 1 for DEO wise results*)

Calculation of PMT Score

Figure 2 below shows that PMT scores were not calculated by NADRA for only 3.8 percent of the total households interviewed. Thus, the PMT score calculation has been compared for the remaining 96.2 percent of the households interviewed.

Figure 2: PMT Score Calculated by NADRA



For all DEOs the proportion of households for which the scores were not calculated was below 5 percent. It was lower for Deloitte and Adv. E-Tech, with 1.31 percent and 2.69 percent households, respectively, for which scores were not calculated by NADRA.

³ For questionnaires following the old format children’s education and number of dependents was taken from the back side and did not depend on the age calculation.

Table 7: DEO Wise PMT Score Calculation

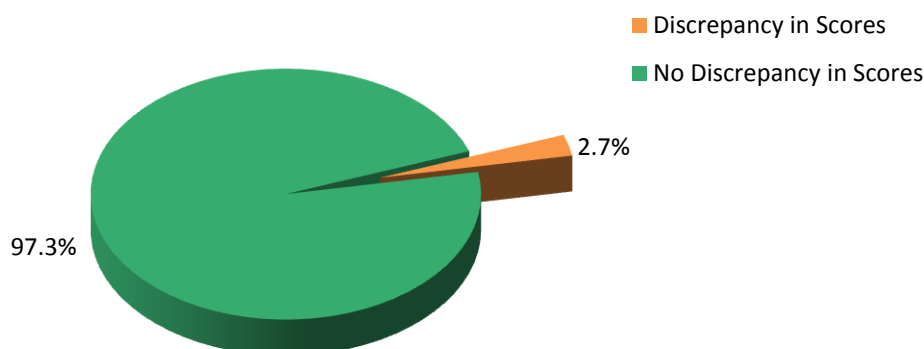
DEO	Number of households surveyed	PMT Score Calculated by NADRA	Percentage of Total Household interviewed	Score not calculated by NADRA	Percentage of Total households interviewed
Adv. E-Tech	135	131	97.03%	4	2.96%
Deloitee	305	301	98.60%	4	1.31%
DPS	490	467	95.30%	23	4.69%
IA	318	311	97.70%	7	2.20%
MYASCO 360	235	225	95.70%	10	4.25%
NCBMS	81	78	96.30%	3	3.70%
NIFT	662	634	95.70%	28	4.20%
Systems	503	479	95.20%	24	4.78%
Others	53	51	96.20%	2	3.78%
Overall	2782	2677	96.20%	105	3.78%

NADRA does not calculate the PMT scores for households that are marked as empty, annulled or discrepant. As per NADRA a discrepant household is defined as a household for which there are enumeration errors in the questionnaire, for example, a response was not selected for one or more questions or multiple responses were chosen for a single response question. In this case it is not possible to determine the true answer. Thus, the household is marked as discrepant and the score is not calculated. As shown in Table 7, NADRA did not calculate the PMT score of 105 or 3.78 percent of the surveyed households.

PMT Score Discrepancy

IDS was left with 2,677 households for PMT score comparison. Of these households, 2.7 percent had a discrepancy in PMT score, i.e. the scores calculated by IDS did not match the scores calculated by NADRA. The PMT scores matched for the remaining 97.3 percent.

Figure 2: Score Discrepancy



DEOs with higher errors in data entry had higher score discrepancy. However, despite no errors in the data entry conducted by the DEOs there may be households with score discrepancy. Such cases are analyzed later.

Table 8 shows the score discrepancies for each DEO. NCBMS and Adv. E-Tech had higher proportions of discrepancy i.e. 7.7 percent and 6.1 percent 5.8 percent and 5.9 percent of their respective samples. Additionally, there was a variation in the scores of 5.9 percent and 5.8 percent of the Others and MYASCO 360 sample, respectively, while for the remaining DEOs this discrepancy existed for less than 4 percent households.

Table 8: PMT score Discrepancy

DEO	Number of Households with Scores Calculated by NADRA	Households with Discrepant Score	Households with Discrepant Score (Percentage)	Households with no Discrepancy in Score	Households with no Discrepancy in Score (Percentage)
Adv. E-Tech	131	8	6.1%	123	93.9%
Deloitte	301	10	3.3%	291	96.7%
DPS	467	8	1.7%	459	98.3%
IA	311	6	1.9%	305	98.1%
MYASCO 360	225	13	5.8%	212	94.2%
NCBMS	78	6	7.7%	72	92.3%
NIFT	634	14	2.2%	620	97.8%
Systems	479	5	1.0%	474	99.0%
Others	51	3	5.9%	48	94.1%
Total	2677	73	2.7%	2604	97.3%

The degree of discrepancy in score varies for the 2.7 percent households with differences in scores calculated by NADRA and IDS. Figure 5 summarizes differences in the two scores. The smallest range of difference was of 0-2.99, which was the second most common margin of error, with 38.9 percent of the score discrepant households falling in this range. The most common range of difference was 5-7.99, with the difference of 43.1 percent of the score discrepant households within this range. Remaining 12.5 percent had a difference of 3 to 4.99 score points, while only 4.2 percent had a difference in score of more than 11 points.

Figure 3: Score Difference Range

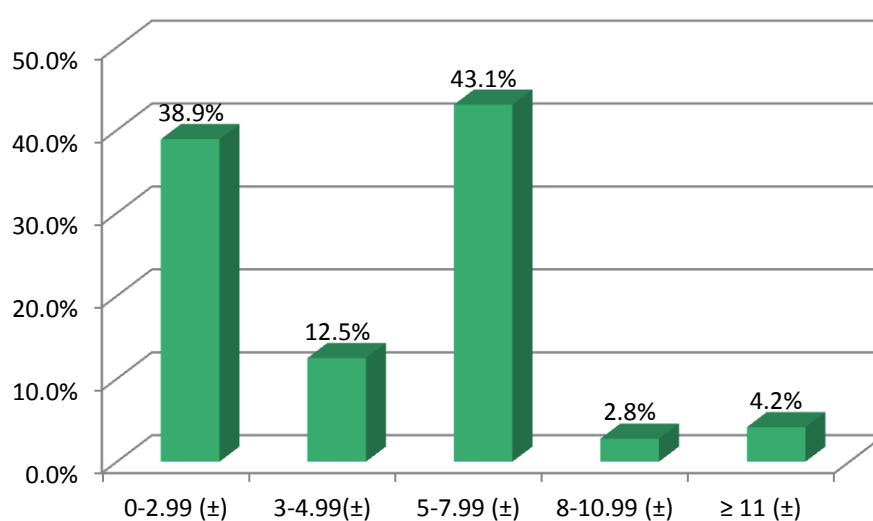


Table 9 reports the differences in the scores calculated by NADRA and IDS across the nine DEOs.

Table 9: DEO wise Difference in Score Range

	Adv. E-Tech	Deloitte	DPS	IA	MYASC O 360	NCBMS	NIFT	Systems	Others
0-2.99 (±)	25.0%	60.0%	50.0%	16.7%	38.5%	33.3%	28.6%	20.0%	100.0%
3-4.99(±)	12.5%	0.0%	25.0%	16.7%	15.4%	0.0%	7.1%	40.0%	0.0%
5-7.99 (±)	62.5%	30.0%	25.0%	66.7%	46.2%	33.3%	57.1%	20.0%	0.0%
8-10.99 (±)	0.0%	10.0%	0.0%	0.0%	0.0%	16.7%	0.0%	0.0%	0.0%
≥ 11 (±)	0.0%	0.0%	0.0%	0.0%	0.0%	16.7%	7.1%	20.0%	0.0%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%

Discrepancy in PMT Scores of Non-Discrepant Households

Out of the total sample of 2,782 households, 94.7 percent or 2,635 households were non-discrepant households. Table 10 shows the variation in the scores of these households. Households that have their data entered correctly should not have a variation in the two scores. However, there was only one case for which the scores did not match.

Table 10: Non-discrepant households with Discrepancy in PMT scores

DEO	Non-Discrepant Households	Score Calculation by NADRA of Non-Discrepant Households	Non-Discrepant Households with difference in Score	Percentage
Adv. E-Tech	118	116	0	0.00%
Deloitte	291	288	0	0.00%
DPS	466	446	0	0.00%
IA	303	299	0	0.00%
MYASCO 360	214	204	0	0.00%
NCBMS	70	67	0	0.00%
NIFT	633	611	0	0.00%
Systems	489	467	0	0.00%
Others	51	49	1	2.04%
Total	2635	2547	1	0.04%

An analysis revealed that the interview of this household was administered on the questionnaires designed according to the old format. For such households the PMT score was calculated up to two decimal places by NADRA. Hence, IDS calculated and compared the score up to two decimal places for this particular household. Consequently, the difference in score was only 0.01 score points. This difference is negligible and could be attributed to variations in rounding off by NADRA and IDS.

Findings

- Despite the numerous safeguards such as the double entry system, inbuilt checks and rigorous monitoring systems, used by the DEOs, data entry errors exist
- Discrepancy in data entry was found to be 5.3 percent in the sample of 2,782 households
- For all DEOs except Systems the percentage of this error was more than 3 percent. It was higher for NCBMS and Adv.E-Tech, i.e. 13.6 percent and 12.6 percent of their respective samples with errors in data entry.
- Data entry errors were only in 1 question for 89.1 percent of the total discrepant households
- Two questions in which discrepancy is larger are: 1) Number of dependents – 61.9 percent and 2) Children’s education – 12.9 percent
- NADRA for reasons specified earlier has been unable to calculate the PMT score of 3.8 percent of the sample households.
- Of the remaining 96.2 percent (2,677) households whose PMT score was calculated, 97.3 percent of the PMT scores calculated by NADRA and IDS matched. In case of 2.7 percent or 73 households the PMT Score did not match. This is primarily because of the 5.3 percent discrepant households where data entry errors were committed.
- Of the 2,635 non-discrepant households, the PMT scores did not match for only one household.
- IDS calculated and compared the score up to two decimal places for this particular household. Consequently, the difference in score was only 0.01 score points. This difference is negligible and could be attributed to variations in rounding off by NADRA and IDS.
- The discrepancy in data entry and PMT score calculation is summarized below

Table 11: Difference in data entry for the DEOs

DEO	Errors in Data Entry	PMT Score not Calculated by NADRA	Discrepancy in PMT Score
Adv. E-Tech	12.6%	3.0%	6.1%
Deloitte	4.6%	1.3%	3.3%
DPS	4.9%	4.7%	1.7%
IA	4.7%	2.2%	1.9%
MYASCO 360	8.9%	4.3%	5.8%
NCBMS	13.6%	3.7%	7.7%
NIFT	4.4%	4.2%	2.2%
Systems	2.8%	4.8%	1.0%
Others	3.8%	3.8%	5.9%
Overall	5.3%	3.8%	2.7%

Conclusion

The seventh phase of Data Entry Spot Check was carried out to give a detailed outline of the quality of the data entry in the DEO's that were mentioned earlier. This analysis shows us that Systems had error below 3 percent, which made them the DEO's with the lowest number of errors. Consequently, they also had the lowest difference in PMT scores. On the other hand, Adv. E-Tech and NCBMS showed the greatest errors in data entry as well as the greatest difference in PMT scores.

Annex 1: DEO wise discrepancy in questions

	Adv. E-Tech	Deloitte	DPS	IA	MYASCO 360	NCBMS	NIFT	Others	Systems
Discrepancy in Number of Dependents	9	5	16	10	12	6	10	0	8
Discrepancy in Children Education	4	3	5	3	6	2	2	0	4
Discrepancy in Air Cooler Ownership	0	0	1	0	1	0	5	0	1
Discrepancy in Heater Ownership	0	1	1	1	0	1	2	0	1
Discrepancy in TV Ownership	2	1	0	0	2	0	2	0	0
Discrepancy in Cooking Stove Ownership	1	0	0	0	0	0	4	0	0
Discrepancy in Tractor Ownership	0	0	1	0	0	0	3	0	1
Discrepancy in Cow Ownership	1	2	0	0	0	1	1	0	0
Discrepancy in Microwave Oven Ownership	0	0	0	1	0	0	2	0	1
Discrepancy in Motorcycle Ownership	1	0	0	1	0	0	1	0	0
Discrepancy in Land Area	0	0	0	1	0	0	1	0	1
Discrepancy in Sheep Ownership	1	0	0	0	0	1	0	1	0
Discrepancy in Buffalo Ownership	1	0	1	0	0	0	1	0	0
Discrepancy in Number of Household Members	0	0	0	0	0	0	1	0	1
Discrepancy in Air Conditioner Ownership	0	1	0	0	0	0	1	0	0
Discrepancy in Scooter Ownership	0	0	0	0	1	0	1	0	0
Discrepancy in Household Head Education	0	1	0	0	0	0	0	0	0
Discrepancy in Number of Rooms	0	0	0	0	0	1	0	0	0
Discrepancy in Type of Toilet	0	1	0	0	0	0	0	0	0
Discrepancy in Refrigerator Ownership	0	0	0	0	0	0	1	0	0
Discrepancy in Washing Machine Ownership	0	0	0	0	1	0	0	0	0
Discrepancy in Bull Ownership	0	0	0	0	0	0	1	0	0
Discrepancy in Land Unit	0	0	0	0	0	0	1	0	0
Discrepancy in Goat Ownership	0	0	0	0	0	0	0	1	0