DATA QUALITY ASSESSMENT OF PERFORMANCE MONITORING PLAN INDICATORS

Prepared for:
USAID/Macedonia

Prepared by
Hap Carr, Team Leader
Nikolina Kenig
Ilija Todoroski
Anica Dragovic

Checchi and Company
Consulting/Louis Berger
Joint Venture

1899 L Street, NW
Suite 800
Washington, D.C. 20036

Contract No. AEP-I-824-00-00022-00
July 2004

I INTRODUCTION

A. Purpose of the Data Quality Assessment

The purpose of this assessment was "to determine whether USAID Macedonia's principal performance indicators are satisfactory in terms of meeting the criteria for quality as outlined in the USAID Automated Directives System (ADS) and the TIPS guidance provided by the Center for Development Information and Evaluation (CDIE)." (See Attachment A for the Scope of Work.)

The Contractor was specifically given three tasks:

Task (1) Review the selection of indicators in terms of whether they meet the following criteria (as defined in the ADS) –

- Objectivity;
- Practicality; and
- Adequacy.

Task (2) Review the criteria (as defined in the ADS) for collecting quality performance data –

- Validity;
- Measurement error;
- Representativeness;
- Reliability and
- Timeliness.

Task (3) Review the documents that record the data and whether there is a detailed specification for each indicator on the following dimensions:

- A comprehensive operational definition and precise unit of measurement;
- A detailed description of the data source, methods for data collection and the frequency of collection.

The Scope of Work also required the contractor to recruit and employ two Macedonian research specialists to form part of the DQA Team. The following task was included in the Scope of Work's section on deliverables which specified:

> "The contract will be carried out in such a manner that the US Team Leader will impart systematically and fully an approach and methodology for assessing data quality so that the Macedonian team members will be able in the future to carry out such assessments."

Upon arrival a fifth task was added. The Team Leader was requested to provide advice to the Mission and each SO team concerning which indicators should be included in the Mission's Annual Report for 2004.

B. Format of the Data Quality Assessment

Section I, the introduction, sets out the purposes to be achieved and tasks to be carried out in the DQA and presents the universe of indicators to be assessed along with the results they are intended to measure.

Section II contains the Summary Performance Indicator Quality Assessments. Here, the methodology for the DQA is described followed by summary findings and recommendations regarding the quality of each indicator. The summary assessment

is given for each indicator against the seven selected DQA criteria listed above. An eighth category of "criteria" is also presented regarding the adequacy of indicator definition in response to Task (3) in the Scope of Work. The recommendations are specifically presented as remedial steps that should be taken in order to improve the quality of indicators where weaknesses may have been found. When practical remedial steps do not appear to the DQA Team then the weakness is presented so that the SO Team can be aware of the problem when interpreting performance monitoring reports in the future.

Section III presents the indicators which are recommended for inclusion in the Mission's upcoming Annual Report.

Section IV addresses the training of Macedonian DQA specialists. It contains a brief description of the training activities, both workshop and on-the-job, which were carried out. An assessment of the process and "how it worked" is given, including the quality of learning obtained by the Macedonian DQA team members, any additional training they might require and how they can be of use to the Mission in the future.

Attachments include the following:

- Attachment A is the formal Scope of Work which defined the assignment.
- Attachment B contains a listing of the persons interviewed, both in the Mission and among the Implementing Partners (IP).
- Attachment C contains perhaps the most important part of the assessment – Data Quality Assessment Checklists for each indicator with detailed observations about the seven DQA criteria focused on in this assessment.
- Attachment D contains the Resumes of the Macedonian Research Specialists trained in the DQA methods.

C. Performance Indicators Covered in the Data Quality Assessment

The indicators which were selected by the Mission for assessment were all those which appear in the Performance Management Plans (PMP) for each of the three Strategic Objectives for USAID Macedonia. They are shown below with the results they are intended to measure. The indicators shown are expressed as they appear in the PMPs. The PMPs are in a matrix format showing the following categories of information for each:

- Definition and unit of measurement;
- Data source and collection method;
- Baselines; and
- Targets.

Result	Indicator
SO 1.3: Accelerated Development and Growth of the Private Sector	(a) Total full time employment in private sector firms
	(b) Private sector employment as a percentage of total employment
IR 1.3.1: Bank and non-bank financial institutions strengthened	(a) Bank lending relative to bank capacity to lend
	(b) Securities markets and non- bank financial institutions
	(c) Market turnover
IR 1.3.2: Private sector firms more competitive	(a) Total exports
	(b) Foreign direct investment
	(c) Total exports of sectors assisted by USAID
IR 1.3.3: Enabling environment for investment improved	(a) Speed of business registration
	(b) Shareholder awareness of their rights
	(c) Additional taxpayers
	(d) Compliance to WTO requirements
SO 2.0: More Legitimate Democratic Institutions	(a) Public perception of respect for democratic values
	(b) Public perception of effectiveness of key government institutions
IR 2.0 1: Increased citizen participation in political and social decision making	(a) Public participation in political activities
	(b) Continuing initiatives at the local level
	(c) Women's participation in community level decision making
	(d) NGO Sustainability Index
IR 2.0 1.1: Citizen attitudes towards democratic processes and practices improved	(a) Public trust in NGOs
IR 2.0.1.2: Improved opportunities for citizen participation	(a) Public participation in civil society organizations
IR 2.0.1.3: Improved citizen access to objective, unbiased information	No Indicator to assess
IR 2.0.2: Adherence to rule of law	(a) Public perception of the effectiveness of courts, both nationally and in the pilot court areas
	(b) Freedom House "Rule of Law" rating
SO 2.0: More Legitimate Democratic Institutions (cont.)	
IR 2.0.2.1: More effective judiciary	(a) Civil case backlog reduction in pilot courts
	(b) User satisfaction with courts, in pilot courts
	(c) Changes in the performance of the justice system
IR 2.0.2.2: More effective legal personnel	(a) User satisfaction with judges, prosecutors and attorneys in pilot courts
	(b) Changes in the performance and ethics of non-judicial professions
IR 2.0 2.3: Laws, regulations and policies that support market-based economy adopted	(a) Key laws and implementing regulations adopted
IR 2.0.3: More effective, responsive and accountable local government	(a) Public trust in local government
IR 2.0.3.1: Local governments have increased responsibility and financial resources	No Indicator to assess
IR 2.0.3.2: Effective relationship between citizens and local government exits	(a) Public participation in local government decision making

Result	Indicator
IR 2.0.3.3: Municipal associations satisfactorily serve the interests of their members	No Indicator to assess
IR 2.0.3.4: Local governments improve the management of municipal services	No Indicator to assess
IR 2.0.4: Increased confidence in government institutions and political processes	(a) Trust in Parliament
IR 2.0.4.1: Parliament functions in a more effective and visibly representative manner	No indicator to assess
IR 2.0.4.2: Degree to which elections are free and fair improved	(a) Free and fair elections held
IR 2.0.4.3: Political party transformed to a platform and performance basis	No Indicator to assess
IR 2.0.4.4: Communication between elements of government and citizens improved	No Indicator to assess
SO 3.4: Mitigate Adverse Social Impact of the Transition to Market-Based Democracies	Indicators being developed and not assessed
IR 3.4.1: Improved quality and relevance of instruction at primary and secondary schools	(a) Increased number of teachers using improved methodology
	(b) Increased number of career centers
IR 3.4.2: Support a training system for professionalization of school directors	(a) Increased number of school directors complete certification training
IR 3.4.3: Increased access to education	(a) Number of students enrolled in Southeastern Europe University
SO 3.4: Mitigate Adverse Social Impact of the Transition to Market-Based Democracies (cont.)	
IR 3.4.4: Local economic development	(a) Sustained business growth at local level: Assisted members in sheep, fruits and vegetables and assisted artisan members
	(b) Internet users per 10,000 people
	(c) SMEs use of ICT to improve their competitiveness

II SUMMARY PERFORMANCE INDICATOR QUALITY ASSESSMENT

A. Introduction

1. Methodology of the Data Quality Assessment

USAID Macedonia contracted one Performance Monitoring Specialist, Harry Carr as Team Leader, from the Checchi and Co. Consulting/Louis Berger Joint Venture, to carry out a Data Quality Assessment (DQA) covering its three Strategic Objectives (SO). The Scope of Work stipulated that two Macedonian socio-economic research specialists be contracted as well to be trained in DQA methods. In the end three such Macedonian research specialists – Nikolina Kenig, Ilija Todorovski and Anica Dragovic – were contracted. The DQA Team Leader spent a total of three weeks in Macedonia between June 14 and July 6 and was joined by the Macedonian team members between June 18 and July 5.

A total of one and half days were spent in the beginning of the assessment in training the Macedonian research specialists in DQA methods. The role of the Macedonian team members was to "learn on the job" by accompanying the Team Leader on field trips and interviews. During the assessment each Macedonian team member chose several indicators to focus on.

The overall purpose of the DQA has been described above in Section I. The Team interviewed USAID officials at all levels related to each SO as well as the staffs of all Implementing Partners (IP) in Skopje. (See Attachment B for a listing of persons interviewed.) This interviewing continued until June 29. The Team Leader spent a total of two days in the field observing SO activities and facilities, interviewing IP field staff, reviewing field documentation and interviewing activity beneficiaries. The DQA Team then spent five days synthesizing its findings and recommendations, writing them up, and presenting them to USAID Macedonia officials.

The DQA Team focused on a total of thirty performance indicators that were in the three SO Performance Management Plan (PM). (See Sub-Section C of the Introductory Section I above for a listing.)

Detailed observations, findings and recommendation about the quality of each performance indicator over six data quality assessment criteria were recorded in Data Quality Assessment Checklists, attached in Attachment C. These criteria for judging an indicator's quality, prescribed in ADS 203.3.3 and included in the Scope of Work are: validity, reliability, timeliness, objectivity, practicality, and adequacy. Included in the validity checklists are sub-components of measurement error and representativeness called for in the DQA Scope of work. Several of these criteria or sub-criteria refer to indicators that are measured quantitatively through integral scaled data and/or indicators that use sampling methods to gather data. Where criteria or sub-criteria did not apply, it is left blank with a "not applicable" note.

Summary findings and recommendations are given below.

B. SO 1.3: Accelerated Development and Growth of the Private Sector

1. SO Level Indicators

(a) Total full time employment in private sector firms

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- Defined as "number of persons employed, in thousands, disaggregated by gender and ethnicity", this is a valid and direct measure of the Strategic Objective. The process is quite rigorous and consistently applied each year. The Government Statistics Office publishes two calculations for this variable: One as reported by the Government Pensions Office which is generally regarded as under-reporting employment and this as taken through an annual labor force survey.

Recommendations for improvement:

1. As important an indicator as this is for SO 1.3, the SO team should continue to monitor and understand the procedures used in the Labor Force Survey possibly by directly observing data gathering, transmission and calculation.

Reliability

Comment of the DQA analyst:
The instruments are applied and data obtained by the State Statistical Office and analyzed. All information dealing with data gathering, field work, quality of control for this DQA were obtained through interviews with the official persons in charge for the Labor Force Survey employed at the State Statistical Office. In the introduction of the Survey it is said that the Labor Force Survey is conducted in accordance to the methodological recommendation of the International Labor Organization (ILO) ratified on the 13 International Conference of Labor Statistician and the recommendation of the European Statistical Bureau (Eurostat). We would like to convey the statement of the interviewees that the Survey is designed and implemented according to the highest international standards.

Timeliness

Findings:
- Until the year 2003 the data were collected once a year throughout the whole territory of the country and included approximately 1.5% of the total number of households. The results were published a month and a half after the data gathering. Since the beginning of the year 2004 there is an attempt to gather data more frequently through four surveys taken during the year, on quarterly basis. The coverage of each survey is 5,000 households, that means all together in a course of a year 20,000 households will be surveyed. The first and third surveys cover the identical population as do the second and the fourth surveys. Beginning in 2005 the results will be published once a year, a month and a half after the fourth survey.

Recommendations for improvement:
1. It would be more convenient for policy makers and other stakeholders if the Labor Force Survey were published on semi-annual basis where the results covering the first and second survey will be preliminary ones and those of the fourth and final survey the definite ones.

(b) Private sector employment as a percentage of total employment

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Recommendations for improvement:
1. As important an indicator as this is for SO 1.3, the SO team should continue to monitor and understand the procedures used in the Labor Force Survey possibly by directly observing data gathering, transmission and calculation.

Timeliness

Findings:
- Until the year 2003 the data were collected once a year throughout the whole territory of the country and included approximately 1.5% of the total number of households. The results were published a month and a half after the data gathering. Since the beginning of the year 2004 there is an attempt to gather data more frequently through four surveys taken during the year, on quarterly basis. The coverage of each survey is 5,000 households, that means all together in a course of a year 20,000 households will be surveyed. The first and third surveys cover the identical population as do the second and the fourth surveys. Beginning in 2005 the results will be published once a year, a month and a half after the fourth survey.

Recommendations for improvement:
1. It would be more convenient for policy makers and other stakeholders if the Labor Force Survey were published on semi-annual basis where the results covering the first and second survey will be preliminary ones and those of the fourth and final survey the definite ones.

2. IR 1.3.1: Bank and non-bank financial institutions strengthened

a. Bank lending relative to bank capacity to lend

The DQA found nothing of significance to report on or about which to make recommendations for improvement regarding this indicator.

b. Securities markets and non- bank financial institutions

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- This indicator is taken from the European Bank for Reconstruction and Development (EBRD) annual report. The DQA Team attempted to study the definitions, data collection methods and calculations used by the EBRD only to learn that all data used were themselves gathered from secondary sources, complied and included in the report from London. Thus, no DQA, through interviews, analysis of definitions and data gathering methods, or direct observation, was done on this indicator.

Recommendations for improvement:

1. The SO Team should find a way to learn more about how these data are collected and calculated as well as the precise definitions that are used in order to better interpret the data.

c. Market turnover

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Definition

Findings:

- As this indicator is currently presented in the PMP there are two complicating elements:
    a. The Baseline is shown in US dollars. Targets and actuals for 2001 and 2002 are shown in US dollars. Targets for 2003, 2004 and 2005 are shown in Denars.
    b. There are five categories of transactions which are tracked by the project and are itemized for 2003 along with the aggregate value of transactions. It is not clear if they were also tracked in earlier years and the itemized values do not add up to the aggregate value.

Recommendations:

1. Denars should be converted, through some generally acceptable calculation and timeliness, to US dollars for the sake of consistency and comparability with the baseline.
2. There is no compelling need, at this sub-IR level, to itemize the categories of transactions. If the SO team wants to show them at all they should be shown for each year, including their itemized targets. Also they should add up correctly.

Validity

Findings:
- Although the indicator is a valid measure of the result, it is an input indicator and it is not a very direct measure of the result – increased capacity of the Macedonian Stock Exchange. Moreover, although it measures the volume of transactions handled it does not measure the quality with which those transactions are handled or the efficiency, perhaps a more valid measure. The indicator measures the volume of transactions which is more validly a measure of the health of the economy and financial markets. The fact that it has gone up in the last three or for years does not necessarily mean that the capacity of the Macedonia Stock Exchange has commensurately increased.

Recommendations for improvement:
1. The SO team should strongly consider measuring the efficiency of the Macedonian Stock Exchange, thus incorporating the costs that it requires to process transactions; a much more direct measure of the intended result.

3. IR 1.3.2: Private sector firms more competitive

a. Total exports

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:
- The extent to which changes in this indicator can be attributed to USAID and its interventions is very limited. Relevant USAID activities involve only five "clusters": cheese and sheep dairy products; lamb; ICT; tourism and fruits and vegetables. Of these there is no program intention to increase the export of fruits and vegetables. The tourism sector is not included in "total exports". The SO team has not yet figured precisely which of the hundreds of categories of ICT goods which might be exported are considered as its program's targets. Finally, the figures on lamb exports are unreliable. Interventions in the wine cluster have not yet started. Only the export of cheese and sheep dairy products on a national level accurately reflect exports over which USAID/Macedonia can claim attribution.

Recommendations for improvement:
1. An indicator closer to the interventions that SO 1.3 carries out should be identified and should be added to the array of indicators for IR 1.3.2.
2. The SO Team should consider dropping this indicator.

Practicality

Findings:
- All figures in the Bulletin are presented in US$. Since the dollar fluctuates relative to the Euro, it is very difficult to perceive the real trends in exports.

Recommendations for improvement:
1. The figures should be presented in US$ adjusted to the Euro.

b. Foreign direct investment

The indicator *foreign direct investment* in Macedonia would seem to be a better measure of the SO 1.3 Team's performance against IR 1.3.3 – "Enabling

environment for investment improved" than against IR 1.3.2 – "Private sector firms more competitive". Although it is not for the DQA Team to make such a recommendation, the SO 1.3 Team should consider moving this indicator to IR 1.3.3. Otherwise, only the following data quality assessment criterion that applies to this indicator.

Practicality

Findings:
- All figures in the Bulletin are presented in US$. Since the dollar fluctuates relative to the Euro, it is very difficult to perceive the real trends in investment.

Recommendations for improvement:
2. The figures should be presented in US$ adjusted to the Euro.

c. Total exports of sectors assisted by USAID

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Definition

Findings:
- The indicator is in fact several indicators. The exports of each cluster is measured and reported separately. Therefore there is no unique measure of the indicator. There is no global measure or "total" reported. Moreover, the tourism cluster is not measured in terms of exports at all.

Recommendations for improvement:
1. Find a measure for the tourism sector that is comparable, even if reported separately, with the exports measure adopted for the other cluster.
2. If there is a way to aggregate the remaining cluster sector exports that should be done. If this doesn't make economic sense, then they should be simply reported separately. However, because of the attribution problem noted above, if they are reported separately there must also be some indication of the extent to which the cluster entities cover the reported sector exports.

Validity

Findings:
- The validly of this indicator is diminished, in some cases significantly because of the attribution problem: USAID activities work only with selected entities in the "assisted sectors" referred to in the indicator while in most cases USAID's activities do not cover the overall sectors assisted. In the cheese sector there is identify between the sector and the USAID assisted entities in it. But this is less true for lamb exports, much less true for IT exports and less still when it comes to fruits and vegetables.
- The fruits and vegetables IP does even try to measure the sector for this indicator, rather it reports exports of assisted entities.

Recommendations for improvement:
1. Regarding the coverage issue – The SO team should make a judgment about whether to report overall exports of USAID assisted sectors or exports of USAID assisted entities.

Timeliness

Findings:
- The data reported do not correct for inflation or, most significantly for exports, currency fluctuations (US dollar to the Euro) that significantly affected the reported exports.

Recommendations for improvement:
1. The SO team, together with the IP should figure a way to adjust export values reported in US dollars for currency fluctuations.

Reliability

Findings:
- The data reported do not correct for inflation or, most significantly for exports, currency fluctuations (US dollar to the Euro) that significantly affected the reported exports.

Recommendations for improvement:
1. The SO team, together with the IP should figure a way to adjust export values reported in US dollars for currency fluctuations.

Adequacy

Findings:
- The intended result, i.e. movement in the exports of selected sectors through MCA cluster interventions, is a long term result. The indicator is similarly at too high a level to reflect short-term "on-trackedness".

Recommendations for improvement:
1. The SO team should add indicators more at the level of process milestones and intermediate outputs.

4. IR 1.3.3: Enabling environment for investment improved

d. Speed of business registration

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:
- There is a slight potential for transcription error since the original data are recorded by hand in ledgers.
- The indicator is based on the assumption that if the amount of time is reduced for registering a business then that should reflect the fact that administrative procedures have improved. However, elapsed time can greatly improve due to other causes, including "under-the-table" payment. The IP noted that changes in processing time due to administrative improvement will be small, requiring precise measurement. Putting it another way, if there were not an opportunity, in business registration, for "rent seeking behavior" the time for processing registration would be reduced to a greater extent than improved administration.

Recommendations for improvement:
1. Although the potential for error is not particularly significant, the project CG & CL Project could explore the possibility of doing periodic checks from either the Commercial Registry of Businesses or the Public Revenues Office.
2. The IP carries the time reported to two decimal places in an attempt to capture changes with the necessary precision.

e. Shareholder awareness of their rights

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Finding:

- The current baseline (2003) is from a random sample survey of the general population, taking from it those respondents who are shareholders. For the second year the IP plans to draw a sample of shareholders directly from a listing in the Commercial Register. There are several implications in this methodology change which could cast doubt on the comparability of future measurement of "shareholder awareness" *vis a vis* the baseline measure.
  - There is an assumption that the 139 shareholders drawn from the general population (baseline) survey are representative of the approximate 220,000 registered shareholders, specifically that they have a relatively low "awareness" level (29%, see reliability below).
  - Next year the real shareholder population will be asked the same question. What conclusions can reasonably be drawn from the number measuring "awareness" that is generated from that survey?
- Because of the method for distinguishing verbal responses from very general open-ended question to any of eight specific categories of shareholder rights (see above for more detail), whether a respondent ended up being categorized as knowledgeable about any of the 8 "rights" looked for would have a lot of "float and bounce" to it.

Recommendations for improvement:

1. The SO Team should be very sure that it understands all the implications of the change in survey methods from baseline to Year One iteration and is comfortable with the baseline drawn in 2003.
2. If there is uncertainty, the SO Team should decide to use the Year One measure, generated from the universe of real shareholders, as the new baseline.
3. In order to reduce the problem of the second finding above, the IP must:
   a. Train enumerators well
   b. Have as few enumerators as possible
   c. Be able to control circumstances of questioning to ensure careful attention of both the enumerator and the respondent.
4. A more ideal way to generate this information would be in writing, with respondents being in a controlled environment and given consistent instructions.

Reliability

Findings:

- It is unclear how the baseline reported (29%) for 2003 was calculated. The project COP was unaware and has requested clarification from the HQ research specialist who made the calculation. Although the COP does intend to use the same question for the shareholders surveyed in 2004 it is absolutely necessary that he make the calculation in the same way that the baseline was made.

Recommendations for improvement:

1. The method for calculating "shareholder awareness" from the survey responses must be documented and put in the form of a manual in order to ensure consistent calculation.

Practicality

Findings:

- The DQA Team Leader asked the IP for the costs of running the survey, both direct and from project overhead such as time of project staff, and he was not able to answer. There are certainly design saving, as he mentioned, that will be recognized in future years. Moreover, the survey provides the IP with considerable more data than just this one indicator.

Recommendations for improvement:

1. The SO team should ask the IP to provide an estimate of the costs so that at least it would have an idea of about the practicability of this indicator.

f. Additional taxpayers

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- The indicator is a proxy. It is measuring taxpaying entities (not individuals). The logic is that as more of these are added to the government registry as taxpayers they represent a conscious choice to join the formal economic sector of Macedonia. They would do this, presumably but reasonably, only if the business environment were more attractive and conducive than the informal sector. Although this does not exactly reflect an investment decision, as the result seeks, it reflects probable anticipated investments.

Recommendations for improvement:

1. The SO team should be conscious that this is a proxy and watch for any breakdown of its valid proxy properties. In its reporting, the SO team should note it as a proxy.

Objectivity and Definition

Finding:

- The wording of the indicator is too vague to show what it really is measuring, i.e. taxpaying entities, not individuals.
- The variety of "taxpaying entities" is staggering and ranged from government agencies to casinos and includes such things associations, political parties and municipalities. At the same time it does not include sole proprietors or partnerships. There doesn't seem to be a significant problem with the breadth of the indicator as it is operationally defined. But the range should be explicit to the SO team and in the records.

Recommendations for improvement:

1. Change the wording so that it is clear that this is measuring new tax paying entities, most of which are businesses and all of which pay some form of business related tax, i.e. employee withholding, VAT, excise, profit.
2. Change the measurement to the "cumulative number of new taxpayers as a percentage of those in the baseline year". (See below.)
3. The SO team should consider all the things that are captured (and not captured) in the indicator and decide if this is the operational definition desired.

Total no. of No. of taxpayers	Percent increase from previous year	No. of additional taxpayers	Cumulative no. of New Taxpayers as a % of baseline
100,000		0	
150,000	50.00%	50,000	50.00%
175,000	16.67%	25,000	75.00%
200,000	14.29%	25,000	100.00%

Data that is collected	Data that can be presented	
	Data that are presented	Data that should be presented

g. Compliance to WTO requirements

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Definition

Findings:

- The indicator is in fact four indicators that, even though it is referred to in the PMP as an index, cannot be aggregated to yield one measure.

Recommendations for improvement:

1. See Validity below

Validity

Findings:

- The indicator is not an index, as shown in the PMP, but rather four separate indicators that cannot be aggregated to yield one measure. Four things are tracked:
  - WTO laws enacted
  - Notifications submitted
  - Governmental coordinating mechanism meeting held and
  - Topic analysis reports published as a measure of institutional capacity for trade analysis

Recommendations for improvement:

1. The SO Team, together with the IP, should develop or find an exiting milestone measure of compliance. Possibly looking at the experience of other countries or in WTO methods for performance measurement some form of standardized or generally accepted milestone measure could be found.
2. The IP, Activity Manager and other SO Team members had earlier designed an indicator for this result that was a composite indicator in which the four dimensions of the results described above were indeed combined. They were advised that the composite indicator as developed by them was too complicated. The present DQA Specialist has reviewed the composite and recommends that the SO Team re-consider it as an acceptable measure of the result with a good deal higher quality than the current four-part indicator that cannot be aggregated. The composite indicator is indeed more complicated but reasonably so.

C. SO 2.0: More Legitimate Democratic Institutions

1. SO Level Indicators

a. Public perception of respect for democratic values

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each. A considerable amount of detail is given below for alternative ways of measuring "public perception of rights" since it is such an important indicator.

Reliability

Findings:

- Each year a separate and new contract is awarded for the survey. The basic instruments and methods are carried over with slight modifications

Recommendations for improvement:

1. Reliability might be better assured if one contractor gathers the data in a long run. The SO team should consider allowing a multi-year contract for the survey.

Timing

Findings:

- The timing of the survey, based on the annual reporting schedule, could present problems. The measurement takes place annually in September or October. Although that is a perfect time for field work (weather, most people available at their homes), this period also coincides with elections (Parliamentary in 2002 and local Government in 2004). These events are considered by the DQA team to be a potential source of distortion of perceptions, especially in regards to trust in Parliament. The PMP's reported perception levels give the best insight into this influence: The actuals increased from 27.7% at 2001 to 47.6 % at 2002 (the election year) and then decreased on 37.2% one year later. It is hard to attribute this fluctuation to USAID's activities while it is likely that the higher reported levels could be a function of higher expectations do to upcoming elections and campaign promises.
- Moreover, unexpected events can also influence personal trust, a real example being a government scandal that was revealed at the time of the survey's administration.

Recommendations for improvement:

1. If the Fall is the only suitable time period for data collection, then potentially distorting social and political events (like elections) have to be taken into account when the data are interpreted. In the same way, other uncontrollable events have to be taken into account.

Objectivity

Findings:

- The measurement represents an average percent of respondents who answered "fully respected" or "somewhat respected" to the question "To what extend are these rights and freedoms respected in our country?" The two negative responses ("violated often" and "sometimes violated") are not taken into account.
- In grouping positive responses the way it is currently practiced, the indicator loses its precision for distinguishing between varying levels of positive perceptions and will yield, for example, the same result for these two different response scenarios:
    a) 2% - 5, 30%-4, 5%-2, 63%-1; and
    b) 30% - 5, 2%-4, 50%-2, 18%-1
- The measurement is reduced in its precision. Because of the grouping of responses from a potential interval scale, the variable is reduced to a nominal scaling.

Recommendations for improvement:

1. The SO Team should consider what is lost by not showing the extent to which people feel their rights are being violated, that is the negative side of the picture, and whether that side should be reflected in the indicator.
2. More precise measurement should take into account and distinguish those respondents whose perceptions differ in strength. (See Exhibit A on the following page for a more detailed description.)
3. The SO Team is spending a good deal of money to collect valid and precise measurements of perceptions. The SO Team should consider ways to preserve that precision rather than lose it. (See Exhibit B on page 17 for a possible Index of a Rights Perceptions Index.)

EXHIBIT A
Comparison of Actual and Illustrative Responses
Showing How Different Responses can Yield the Same Measure

Rights	Scenario One (actual 2003 survey data)			Scenario Two (illustrative data)		
	Two Positive Responses	Valid Percent	Aggregate Valid Percent of Positive Responses	Two Positive Responses	Valid Percent	Aggregate Valid Percent of Positive Responses
Press	Fully Respected	10.8	22.3	Fully Respected	5.3	22.3
	Somewhat Respected	11.5		Somewhat Respected	17.0	
Vote	Fully Respected	20.7	35.3	Fully Respected	3.5	35.3
	Somewhat Respected	14.6		Somewhat Respected	31.8	
Speech	Fully Respected	28.6	42.5	Fully Respected	7.7	42.5
	Somewhat Respected	13.9		Somewhat Respected	34.8	
Personal Safety	Fully Respected	10.8	20.0	Fully Respected	2.1	20.0
	Somewhat Respected	9.2		Somewhat Respected	17.9	
Equality	Fully Respected	8.2	14.4	Fully Respected	4.6	14.4
	Somewhat Respected	6.2		Somewhat Respected	9.8	
Health Care	Fully Respected	16.3	30.0	Fully Respected	3.9	30.0
	Somewhat Respected	13.7		Somewhat Respected	26.1	
Religion, etc.	Fully Respected	33.2	49.8	Fully Respected	5.8	49.8
	Somewhat Respected	16.6		Somewhat Respected	44.0	
Healthy Environ.	Fully Respected	12.5	24.8	Fully Respected	8.3	24.8
	Somewhat Respected	12.3		Somewhat Respected	16.5	
Strike	Fully Respected	22.7	38.3	Fully Respected	4.2	38.3
	Somewhat Respected	15.6		Somewhat Respected	34.1	
Pol. Association	Fully Respected	35.6	52.2	Fully Respected	5.7	52.2
	Somewhat Respected	16.6		Somewhat Respected	46.5	
Organize	Fully Respected	34.5	51.4	Fully Respected	3.4	51.4
	Somewhat Respected	16.9		Somewhat Respected	48.0	
Free Enterprise	Fully Respected	22.3	38.2	Fully Respected	6.1	38.2
	Somewhat Respected	15.9		Somewhat Respected	32.1	
Education	Fully Respected	33.6	48.1	Fully Respected	7.3	48.1
	Somewhat Respected	14.5		Somewhat Respected	40.8	
Private Property	Fully Respected	32.9	49.5	Fully Respected	8.3	49.5
	Somewhat Respected	16.6		Somewhat Respected	41.2	
Work	Fully Respected	13.3	19.9	Fully Respected	1.4	19.9
	Somewhat Respected	6.6		Somewhat Respected	18.5	
Social Security	Fully Respected	13.4	22.7	Fully Respected	6.2	22.7
	Somewhat Respected	9.3		Somewhat Respected	16.5	
SO 2.0 (a) Indicator:			34.96			34.96
"Public perception of respect for rights"						

EXHIBIT B
Alternative Methods for Calculating
"Public Perceptions of Respect for Democratic Values"
Including an Index

Rights	Positive and Negative Responses	Score	Valid Percent (from actual '03)	Aggregate Valid Percent of Positive Responses	Aggregate Valid Percent of Negative Responses	Weighted Average Score	Rights Perception Index
To Free Press	Fully Respected	5	10.8	22.3		171.6	42.9
	Somewhat Respected	4	11.5				
	Violated Often	2	25.6		46.0		
	Sometimes Violated	1	20.4				
To Vote	Fully Respected	5	20.7	35.3		217.9	54.5
	Somewhat Respected	4	14.6				
	Violated Often	2	19.4		36.6		
	Sometimes Violated	1	17.2				
To Speech	Fully Respected	5	28.6	42.5		252.1	63.0
	Somewhat Respected	4	13.9				
	Violated Often	2	19.3		34.2		
	Sometimes Violated	1	14.9				
To Personal Safety	Fully Respected	5	10.8	20.0		179.2	44.8
	Somewhat Respected	4	9.2				
	Violated Often	2	34.6		53.8		
	Sometimes Violated	1	19.2				
To Equality	Fully Respected	5	8.2	14.4		172.4	43.1
	Somewhat Respected	4	6.2				
	Violated Often	2	42.7		63.9		
	Sometimes Violated	1	21.2				
To Health Care	Fully Respected	5	16.3	30.0		207.5	51.9
	Somewhat Respected	4	13.7				
	Violated Often	2	25.9		45.3		
	Sometimes Violated	1	19.4				
To Religion, etc.	Fully Respected	5	33.2	49.8		267.7	66.9
	Somewhat Respected	4	16.6				
	Violated Often	2	12.3		23.0		
	Sometimes Violated	1	10.7				
To Healthy Environ.	Fully Respected	5	12.5	24.8		183.6	45.9
	Somewhat Respected	4	12.3				
	Violated Often	2	26.9		45.0		
	Sometimes Violated	1	18.1				
To Strike	Fully Respected	5	22.7	38.3		221.6	55.4
	Somewhat Respected	4	15.6				
	Violated Often	2	16.2		29.5		
	Sometimes Violated	1	13.3				
To Pol. Association	Fully Respected	5	35.6	52.2		276.8	69.2

Rights	Positive and Negative Responses	Score	Valid Percent (from actual '03)	Aggregate Valid Percent of Positive Responses	Aggregate Valid Percent of Negative Responses	Weighted Average Score	Rights Perception Index
	Somewhat Respected	4	16.6				
	Violated Often	2	11.7		20.7		
	Sometimes Violated	1	9.0				
To Organize	Fully Respected	5	34.5	51.4		269.0	67.3
	Somewhat Respected	4	16.9				
	Violated Often	2	10.4		18.5		
	Sometimes Violated	1	8.1				
To Free Enterprise	Fully Respected	5	22.3	38.2		206.7	51.7
	Somewhat Respected	4	15.9				
	Violated Often	2	11.2		20.4		
	Sometimes Violated	1	9.2				
To Education	Fully Respected	5	33.6	48.1		282.5	70.6
	Somewhat Respected	4	14.5				
	Violated Often	2	21.9		34.6		
	Sometimes Violated	1	12.7				
To Private Property	Fully Respected	5	32.9	49.5		273.9	68.5
	Somewhat Respected	4	16.6				
	Violated Often	2	16.1		26.9		
	Sometimes Violated	1	10.8				
To Work	Fully Respected	5	13.3	19.9		209.4	52.4
	Somewhat Respected	4	6.6				
	Violated Often	2	50.1		66.4		
	Sometimes Violated	1	16.3				
To Social Security	Fully Respected	5	13.4	22.7		196.0	49.0
	Somewhat Respected	4	9.3				
	Violated Often	2	36.6		55.2		
	Sometimes Violated	1	18.6				

SO 2.0 (a) Indicator: "Public perception of respect..." 35.0

Alternative indicator: "Public perception of violation of rights" 37.7

Index (un-weighted average) 56.1

Index calculation:

(1) Assume all five response categories are equally probable at 20%

(2) Eliminate the neutral response category

(3) Take the weighted average score over response categories 1 and 2 and 4 and 5

(4) Assume a maximum score of 400 for any given category

(5) Subtract the weighted average score from the maximum possible score of 400 and divide by 4 to bring it to a 100 point index

(6) Subtract that score from 100 to adjust for the direction the index takes to reflect direction of score points

b. Public perception of effectiveness of key government institutions

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

The data are taken from three different questions in the questionnaire:

- For perception of the Parliament, government and local government - Respondents are asked "How effective, do you think are these institutions in solving problems?" They can assess the efficacy by choosing one of four alternatives: a) very effective, b) effective to some extent, c) ineffective to some extent and d) not effective at all. Six institutions are listed but only 3 are taken into account for this indicator (parliament, government, local government).
- For perception of the judiciary. Respondents are asked "Do you think the courts in our country are effective?" They can assess the efficacy by choosing one of four alternatives: a) very effective, b) effective to some extent, c) ineffective to some extent and d) not effective at all. It is not very clear why the question for the courts is pulled out as a separate one.
- The question regarding the NGOs is about usefulness as opposed to efficacy and is structured in an entirely different way from the other questions. It says "To what extent do you consider NGOs … **useful**?" and asks the respondent to make separate judgments for 11 different NGO sectors. The alternative responses vary from "very useful" (5) to "not useful" (1) where intermediate perceptions are simply scored "4", "3", and "2". The final NGO "usefulness" measure is aggregated as an overall average of those responding "very useful" and those marking "4".

Recommendations for improvement:

1. Adjust the questions so that they ask respondents the same thing in the same way (for the 5 different institutions) and provide the same scale for the answers.

Reliability

See "reliability" under C.1.a above

Timing

See "timing" under C.1.a above

Objectivity

Findings:

- The measurement is an average of people who assigned grade 5 or 4 for the efficacy of the Parliament, the Government, the local government and the courts.
- The result for the question for usefulness of the NGOs (Q31) is calculated by adding the percentages of persons who consider them being "very useful" or "useful" (presumably the category called "4") one by one and then dividing the sum by 22 (number of categories). This approach is somewhat different then the previous one.

Recommendations for improvement:

1. Question 31 (NGOs) can be asked in a general form first (not for usefulness but for efficacy), and then, if the information about separate categories and usefulness is needed it can remain as such.

2. IR 2.0.1: Increased citizen participation in political and social decision making

a. Public participation in political activities

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Reliability

See "reliability" under C.1.a above

Timeliness

See "timing" under C.1.a above

Objectivity

Findings:

- The definition of this indicator is "average percent of citizens responding that they have participated in several types of political action". There are 13 possible choices listed (e.g. run for an office, vote in local elections, join strikes etc.) with no open-ended option added. The average is counted by adding up the percentages of respondents who answered positively at each of these options, then divided by the number of alternatives. Thus all forms of participation are counted equally, i.e. as having the same importance.

Recommendations for improvement:

1. Taking into account that the alternatives can be seen not only as bearing a positive vs. bearing a negative connotation, but also as very different in the quality and quantity of personal engagement needed to accomplish them, it would be very useful to assign different weight to different options prior to aggregating them in one average number. For the sake of comparability that should be done for the previous years as well.
2. The DQA Team is recommending that this indicator be included in the Annual Report if the SO Team, together with its implementing partners, can agree on weighting for each form of participation.

b. Continuing initiatives at the local level

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each. Since this indicator is essentially two indicators, with different definitions and data gathering methods, and since the DQA is recommending they be reported separately, two DQA Checklists have been prepared (see Attachment C) and their findings and recommendations are presented separately below.

Definition

Findings:

- The indicator is defined as the "percent of citizens groups that have already organized and implemented activities through CSHI or ISC (two projects with different IPs) and are continuing with the participatory planning and implementation process. Groups are counted if they (a) are seeking assistance from other donors or (b) have organized local self-help contributions for another (follow-on) community activity". One IP reports two numbers against (a) and (b) while the other reports one number aggregated number against both categories (a) and (b). Those three numbers are added together, with each component having equal weight, to yield one aggregated measure.

  Both programs are extremely different in purpose, approach, mechanisms, inputs and types of clients served. The aggregated reported measure tends to overstate progress as practiced by one IP and to understate progress as seen by the other. The single aggregated measure does not effectively measure the USAID program promoting "citizen participation". It simply is not meaningful.

Recommendations for improvement:

1. The progress of each IP should be reported separately. The D/G program itself, particularly as it addresses civil society is currently in transition. The SO team should take this opportunity to consider the two approaches and make a choice between them reflecting the sorts of results it wants to achieve.

For IR 2.0.1 (b) as collected by ISC:

Validity of the indicator as interpreted by ISC:

Findings:

- The "data" reported are qualitative and subjectively arrived at, reasonably so given the nature of the result that is being captured. There are several ways, however, in which the IP (ISC) and SO team are attempting to ensure validity.
  a. Both parties have experimented with different methods for measuring the result and have found themselves in agreement in their dissatisfaction.
  b. The IP has developed a new data gathering method and approach (reported on here) that is an improvement in terms of objectivity and verifiability.
  c. The new monitoring method and approach is more costly than that used before but more certain in its analytical results.

Recommendations for improvement:

1. The SO team and IP should continue using the new approach (Performance Stories) and experiment, as it is, with improving it regarding its measurement of this result;
2. The IP should continue reporting its results against this indicator in three categories as it does now and has been in the past, for the sake of continuity;
3. The IP should document as much as possible all of its reported findings in these categories;
4. The SO team should make a choice between the three categories (definitions of "continuity of initiatives") which is the most appropriate given how the SO team sees the essence of the result and report on that particular category.

Practicality of the indicator as interpreted by ISC:

Findings:

- The data gathering methods are changing now in that an independent assessment is now being contracted to give validation and depth to the results reported. The costs are not ascertained as of yet.

Recommendations for improvement:

1. Additional costs to some magnitude are easily justified in giving the indicator more objectivity and validity.

For IR 2.0.1 (b) as collected by CSHI:

Validity of the indicator as interpreted by CSHI:

Findings:

- The "data" reported are qualitative and subjectively arrived at, reasonably so given the nature of the result that is being captured. There are several ways, however, in which the IP (CSHI) is attempting to ensure validity. In order to be classified as a "continuing initiative" the following things have to happen:
  a. The group which received the CSHI grant should be reflected formally in the "group" which is applying for or otherwise obtaining funding for a second initiative, i.e. it doesn't need to be exactly the same persons but it should be largely so;
  b. The new initiative has gone through a project preparation process similar to that in which the original CSHI group was trained;
  c. There must be a concrete project idea which has been developed by the group; and
  d. The conceptualization has to take some concrete and demonstrable form or documentation.
- This process very much depends on the close oversight of and commitment by project staff in both the M&E project department and the regional representative through on-going visits and reports. As observed in field visits this process appears to be rigorously and objectively carried out. However only two sub-projects out of some 300 were visited.

Recommendations for improvement:

1. The SO team must keep very close contact with the IP and its M&E department, review the field visit reports for projects characterized as having continued initiative, and make site visits to these projects and reach agreement of the projects so classified.

c. Women's participation in community level decision making

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Definition

Findings:

- This indicator is reported on by two implementing partners, as above, and the inputs from each are aggregated.
  a. As of the end of the DQA it was not determined how ISC defines and measures "women participating in community-level decision-making. (There is no DQA Checklist, therefore, prepared for it.)
  b. CSHI defines and measures the indicator people who are in the official (and formalized) role of leader, as specified in the Memorandum of Agreement who are women.

Recommendations for improvement:

1. The SO team should determine how ISC defines and measures women in leadership roles and decide if that definition and form of measurement is consistent with the CSHI method.

Validity

Findings:

- It was not observed nor questioned during the DQA whether and to what extent the CSHI project actually places an emphasis on choosing women to be in leadership positions and has particular interventions or techniques designed to make this happen, i.e. is not clear if this is a conscious objective of the project. It is not included or alluded to in the Project's Mission Statement nor is among five project Program Objectives as is, for example, "promoting ethnic harmony by encouraging collaboration and cooperation among diverse ethnic groups in the planning and implementation of project activities."

Recommendations for improvement:

1. This question should be answered by the SO team. If, indeed, there is no explicit objective for promoting leadership among women then the indicator should be dropped.

d. NGO Sustainability Index

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- There are currently 4,000 NGOs registered in Macedonia. USAID/Macedonia deals with less than 10% of that number. Of the two activities designed to develop NGO sustainability only one works in-depth over time with its constituents but they approximate ten organizations. The other, which works with the majority of USAID assisted NGOs, works with them only during grant-funded project implementation, i.e. the focus is on completing projects, most often of small scale infrastructure nature.
- The broader-gauged interventions, involving capacity building, financial viability, advocacy and service provision, are carried out by the activity referred to above which focuses on only ten NGOs (varies over the reporting period) and these have been chosen because of particular geographic, gender or ethnic criteria.

Recommendations for improvement:

1. The SO Team should focus its measure of sustainability on those institutions with which it works and it should develop a more narrowly focused measure of sustainability than the NGO Sustainability Index.

3. IR 2.0 1.1: Citizen attitudes towards democratic processes and practices improved

a. Public trust in NGOs

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- The degree of trust is measured by asking respondents "Do you trust the following institutions." The question is accompanied with 10 different types of institutions and the respondent has to rate his/her degree of trust on a four- point scale (definitely yes, rather yes, rather no and definitely no). The alternative institutions are read in the same order to all participants in the survey, with the NGO's appearing 8th on the list, after labor unions, parliament, Government of RM etc. There is a theoretical possibility that respondents' sense of trust in NGOs might be influenced by the types of institutions immediately preceding that alternative and that a different ordering of the alternatives could create a different psychological frame of reference for estimating the trust.

Recommendations for improvement:

1. The alternatives should be rotated from one to another interview in all possible combinations, so that the "effect of ordering" is annulled.

Reliability

See "reliability" under C.1.a above

Timeliness

See "timing" under C.1.a above

4. IR 2.0.1.2: Improved opportunities for citizen participation

a. Public participation in civil society organizations

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Reliability

See "reliability" under C.1.a above

Timeliness

See "timing" under C.1.a above

Objectivity

Findings:

- The measurement is based on self-report, which might or might not correspond to the reality. Moreover, it is unknown what sort of "group or organization" the respondent may have joined and whether this is one that the SO Team would want to count. A more objective measure might be counting the number of active members of a representative sample of different interest groups of the sort the SO Team considers important.

Recommendations for improvement:

1. A more objective measure might be counting the number of active members of a representative sample of different interest groups of the sort the SO Team considers important.

5. IR 2.0.1.3: Improved citizen access to objective, unbiased information

a. No indicator to assess

6. IR 2.0.2: Adherence to the Rule of Law Enhanced

a. Public perception of the effectiveness of courts both nationally and in the pilot court areas

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- The measure is the percentage of respondents that find the courts to be "very effective" or "effective to some extent" nationwide when asked a specific question in the USAID Survey of Citizen Attitudes. The indicator does not, however, measure respondents' perceptions of pilot court effectiveness.

Recommendations for improvement:

1. The indicator should be modified so that it does not include pilot court areas.

Reliability

See "reliability" under C.1.a above

Timeliness

See "timing" under C.1.a above

b. Macedonia's Rule of Law rating in the Freedom House Report

The DQA Team did not assess this indicator. Data for it are collected, analyzed and reported in Washington.

7. IR 2.0.2.1: More effective judiciary

a. Civil case backlog reduction in pilot courts

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- There are two parts of the indicator. The program aims to reduce the backlog of cases open for more than one year and for more than three years from baseline levels measured by a survey of cases in 2003 to set target levels. These measures are standard measures used by the EU and the American Bar Association.

Recommendations for improvement:

1. The SO team should note in the PMP and its reports if there is a higher level of priority given to either of these timeframes in order to assist interpretation of the two measures.

b. User satisfaction with courts, in pilot courts

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Definition

Findings:

- The PMP for SO 1.3 (November 10, 2003 version) defines this indicator as a three-part measure of satisfaction as gathered in the Survey of Citizens and Court Users: Overall satisfaction; satisfaction with the competency of the prosecutor; and satisfaction with the competency of the representing attorney. In fact, however, only the first type of satisfaction – overall - is used to measure of IR 2.0 2.1.

Recommendation for improvement:

1. The SO team should correct the PMP.

Validity

Findings:

- On the surface the indicator is a valid measure of one dimension of effectiveness: court user perception of satisfaction. However, there are two serious problems with how the indicator is measured:
    - It is entirely based on people's response to one question asking the respondent to characterize his or her level of agreement with the statement, "Overall, I think the court performed effectively."
    - It does not distinguish users (citizens) from court officials or other workers who may have been in the court the days the survey was taken. All responses are aggregated. Yet the indicator explicitly aims at "user satisfaction."

- It is not certain that the day selected for surveying in each pilot court was representative. Obviously troublesome days such as Fridays are not selected for sampling. But outside of the obvious problem days or periods there was no systematic way of ensuring representativeness, if indeed there is any way.
- The circumstances in which the survey was administered were not conducive to thoughtful responses. People are quickly stopped as they leave court and asked to fill out the short questionnaire. Writing space is provided. Although the process was not observed, it must have been the case that at least at certain times circumstances were chaotic.

Recommendations for improvement:

1. If this is the way user satisfaction is to be measured the IP should develop a bit more of a battery of questions geared towards measuring it then they should be aggregated.
2. There should be some separation-out of respondents who had no particular interest in the effectiveness of the court that day, if not to focus exclusively on specific types of respondents as having more salient opinions than others.
3. The IP should be asked to be as rigorous as possible in the selection of "survey day" to ensure the representativeness of the sample of court users. The IP should also document what sorts of criteria are being used to make the selection.
4. The IP should ensure an environment for people to thoughtfully answer the questions.

Reliability

Findings:

- The IP must make certain that the procedures and circumstances of data collection of each of the seven pilot courts are the same.

Recommendations for improvement:

1. (See validity above)

c. Changes in the performance of the justice system

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Definition

Findings:

- The indicator is a sub-set of the factors measured in the Judicial Reform Index, taking five of them and reporting on them separately. As such, makes it difficult for the SO team to interpret the results, i.e. how to weigh positive performance on any one as compared to negative performance on others and to arrive at some IR level comprehensive judgment.

Recommendations for improvement:

1. The SO team should work closely with the IP to help give the kind of comprehensive interpretation needed for its management concerns.

Validity

Findings:

- The indicator is at such a high level that changes could not very reasonably be attributed to USAID's efforts.
- The level of precision of each of the components of the indicator is measured in only three gradations. Because each component is at such a high level, changes from one year to the next will be so small that they are not likely to be picked up by the indicator components. They will not likely tell whether the program is "on track". They do not serve as a management tool so much as a measure impact.

Recommendations for improvement:

1. The SO team should be aware of this when making claims for the program based on these indicators.
2. The SO team should realize that measurable results under these indicator components will be a long time in appearing and that changes may not be picked up at all.

Practicality

Findings:

- Given the high level measured by these indicator components, the low level of attribution that USAID can reasonably claim for positive changes and the grossness of precision with which the measurement is made (see other criteria statements in the DQA), it could quite reasonably be argued that the costs are inappropriate for what the information yielded the SO team. The indicator is much more valuable at the regional level, to make comparisons between USAID countries.

Recommendations for improvement:

1. The SO Team should understand the costs involved and consider developing lower level, more precise indicators.

8. IR 2.0.2.2: More effective legal personnel

a. User satisfaction with judges, prosecutors and attorneys in pilot courts

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Definition

Findings:

- The PMP for SO 1.3 (November 10, 2003 version) defines this indicator as a three-part measure of satisfaction as gathered in the Survey of Citizens and Court Users: Overall satisfaction; satisfaction with the competency of the prosecutor; and satisfaction with the competency of the representing attorney. In fact, however, it is the combination of the second two types of satisfaction, with the competence of prosecutors and attorneys, that is used to measure of IR 2.0 2.2.

Recommendation for improvement:

1. The SO team should correct the PMP.
2. The IP and SO team should together consider aggregating these two survey question responses by, for example, reporting the percent of all respondents who answer "strongly agree" and "agree" to either of these questions or, as a stricter view of competency, those who answered .

	Total no. of respondents	No. of respondents answering: strongly "agree" and "agree"	Percent of respondents	Combined answering positively to either (ass
The public prosecutor competently represented the interests of the state against the accused	400	87	21.8%	32.5%
The private attorney competently represented the interests of his/her client	400	43	10.8%	

Validity

Findings:

- The indicator does not distinguish users – citizens - from court officials or other workers who may have been in the court the days the survey was taken. All responses are aggregated. Yet the indicator explicitly aims at "user satisfaction".
- It is not certain that the day selected for surveying in each pilot court was representative. Obviously troublesome days such as Fridays are not selected for sampling. But outside of the obvious problem days or periods there was no systematic way of ensuring representativeness, if indeed there is any way.
- The circumstances in which the survey was administered were not conducive to thoughtful responses. People are quickly stopped as they leave court and asked to fill out the short questionnaire. Writing space is provided. Although the process was not observed, it must have been the case that at least at certain times circumstances were chaotic.

Recommendations for improvement:

1. There should be some separation out of respondents who had no particular interest in the effectiveness of the court that day, if not to focus exclusively on specific types of respondents as having more salient opinions than others.
2. The IP should be asked to be as rigorous as possible in the selection of "survey day" to ensure the representativeness of the sample of court users and the IP should document what sorts of criteria are being used to make the selection.
3. The IP should ensure an environment for people to thoughtfully answer the questions.

Reliability

Findings:
- The IP must make certain that the procedures and circumstances of data collection each of the seven pilot courts are the same.

Recommendations for improvement:
1. (See validity above)

b. Changes in the performance and ethics of non-judicial professions

The following data quality assessment criteria apply to this indicator. The same findings and recommendations as were formulated for IR 2.0.2.1 above – *"Changes in the performance of the justice system"* – are given here since the exact same approach and data gathering methods are used for the Legal Profession Reform Index as for the Judicial Reform Index.

Definition

Findings:
- The indicator, like the Judicial Reform Index discussed above, is a sub-set of the factors measured in the Legal Profession Reform Index, taking them and reporting on them separately. As such, it is difficult for the SO team to interpret the results, i.e. how to weigh positive performance on any one as compared to negative performance on others and to arrive at some IR level comprehensive judgment.

Recommendations for improvement:
1. The SO team should work closely with the IP to help give the kind of comprehensive interpretation needed for its management concerns.

Validity

Findings:
- The indicator is at such a high level that changes could not very reasonably be attributed to USAID's efforts.
- The level of precision of each of the components of the indicator is measured in only three gradations. Because each component is at such a high level, changes from one year to the next will be so small that they are not likely to be picked up by the indicator components. They will not likely tell whether the program is "on track". They do not serve as a management tool so much as a measure impact.

Recommendations for improvement:
1. The SO team should be aware of this when making claims for the program based on these indicators.
2. The SO team should realize that measurable results under these indicator components will be a long time in appearing and that changes may not be picked up at all.

Practicality

Findings:

- Given the high level measured by these indicator components, the low level of attribution that USAID can reasonably claim for positive changes and the grossness of precision with which the measurement is made (see other criteria statements in the DQA), it could quite reasonably be argued that the costs are inappropriate for what the information yielded the SO team. The indicator is much more valuable at the regional level, to make comparisons between USAID countries.

Recommendations for improvement:

2. The SO Team should understand the costs involved and consider developing lower level, more precise indicators.

9. IR 2.0.2.3: Laws, regulations and policies that support market-based economy adopted

a. Key laws and implementing regulations adopted

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- The Mission, through two SO teams, is working for the adoption of these laws and regulations. However, the level of attribution to USAID interventions varies from one to another. And in most cases the argument that the laws and regulations would not have been adopted but for USAID action is at least arguable.

Recommendations for improvement:

1. The Mission, through its two SO teams should be prepared to document and demonstrate the extent of which adoption of each law and regulation can be attributed to USAID.

Practicality

Findings:

- Given the length of time to achieve adoption of these laws and regulations and then the months to get that published in the Official Gazette make it a very imprecise measure for "on-trackedness".

Recommendations for improvement:

1. There is not much to recommend. It takes as long as it takes.

10. IR 2.0.3: More Effective, Responsive and Accountable Local Government

a. Public trust in local government in target municipalities

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Reliability

See "reliability" under C.1.a above

Timeliness

See "timing" under C.1.a above

11. IR 2.0.3.1: Local governments have increased responsibility and financial resources

a. No indicator to assess

12. IR 2.0.3.2: Effective relationship between citizens and local government exits

a. Public participation in local government decision making

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:

- Respondents in the nationwide USAID Survey of Citizen Attitudes are asked "Have you attempted to make the self-government solve a community problem". This indicator is being measured by the percentage of respondents who answered "yes" to the question. However, the result that the indicator is supposed to measure is an increase in public participation in local government decision making. This question does not identify what sort of community problem local government was asked to address nor does it show how that request was made or how powerfully it might have been conveyed.
- The SO Team is considering, with this indicator taken from a nationwide survey, that a change in the indicator could be attributed to actions taken USAID. Yet USAID is working in 17 selected municipalities.

Recommendations for improvement:

1. The SO Team should either have several sub-questions added to the nationwide survey trying to get at determining whether the type of intervention a respondent says he/she "attempted to make" qualifies for the type of citizen involvement in decision making that the Team is aiming at increasing or the SO Team should formulate a new indicator.
2. If the question remains the source of this indicator, qualified as recommended above, then it should envelope only the 17 USAID assisted municipalities.

Reliability

See "reliability" under C.1.a above

Timeliness

See "timing" under C.1.a above

13. IR 2.0.3.3: Municipal associations satisfactorily serve the interests of their members

a. No indicator to assess

14. IR 2.0.3.4: Local governments improve the management of municipal services

a. No indicator to assess

15. IR 2.0.4: Increased confidence in government institutions and political processes

a. Trust in Parliament

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Reliability
See "reliability" under C.1.a above

Timeliness
See "timing" under C.1.a above

16. IR 2.0.4.1: Parliament functions in a more effective and visibly representative manner

a. No indicator to assess

17. IR 2.0.4.2: Degree to which elections are free and fair improved

a. Free and fair elections held

There was not time during the Data Quality Assessment to verify data gathering methods or analyze the validity and adequacy of the indicator. Since data gathering involves using the OSCE observers report and judgment, there really are no issues as regards data gathering. The major issue to be analyzed is the issue of attribution: to what extent can the USAID Mission and the SO 2.0 team claim success for free and fair elections during the next round of voting.

18. IR 2.0.4.3: Political party transformed to a platform and performance basis

a. No indicator to assess

19. IR 2.0.4.4: Communication between elements of government and citizens improved

a. No indicator to assess

D. SO 3.4: Mitigate Social Impact of the Transition to Market-Based Democracies

1. SO Level Indicators

There currently are no SO level indicators for SO 3.4. The former indicator *"Increased number of youth better prepared for employment through education programs"* is being revised by the Implementing Partner for the Secondary Education Activity and baseline data have been taken through two surveys, one in March/April and one during the DQA in June, 2004. The DQA team interviewed the Research Scientist of the IP on two occasions and observed one part of the data entry procedure. The IP is in the process of developing five indices which, together, will more precisely measure the employability of youth as a result of teacher training in new methods. The five indices are:

- Instructional practice;
- Career preparation;
- Computer skills;
- Computer access; and
- Problem solving skills.

The SO Team Leader decided that it was too early to assess the quality of these indices as they are not yet operational.

2. IR 3.4.1: Improved quality and relevance of instruction at primary and secondary schools

a. Increased number of teachers using improved methodology

This indicator was originally formulated for one activity, "Creative Teaching and Learning" (FOSIM being the implementer). More recently two activities have come on-stream and the SO 3.4 Team thought of including those results under this indicator. However, the SO Team has decided that in the future the Secondary Education Activity will not be reporting under it. Finally, the third and newest activity, which is implemented by EDC and will focus on setting up ICT labs in schools, will also train teachers to teach ICT. As of the assessment the EDC activity remained part of this indicator. The FOSIM and EDC activities are sufficiently different as to warrant a separate assessment of each as it contributes to this indicator.

i. *Increased number of teachers using improved methodology as per the FOSIM activity*

Validity

Findings:

- The indicator is measured as an average percent of teachers who answered (in a self-reporting questionnaire) that they either a) often or b) every day use the creative techniques. There is a potential threat that the answers to that question represent teachers' unconscious tendency to present themselves in a way that is expected by the questioner. The validity of the indicator is controlled, in a way, by comparing the percent of teachers reporting that they use the creative techniques with the proportion of students choosing the alternatives a) several times a week and b) once a week offered with the question *"Are most of the teachers using new teaching techniques?"* (It should be noted however, that there is a discrepancy in the estimation of these two groups of respondents; approx. 40% of students think that teachers rarely use new techniques, vs. only one fourth of the teachers who says that the creative methods are used rarely.)
- The indicator is defined as "trained teachers in selected schools demonstrated improved teaching methodology and skills", but no measures have been developed to assess their actual employment of these techniques. It is only the self-report that is taken into account.

Recommendations for improvement:

1. The validity of the indicator could be improved by using, along with surveys administered to the teachers, more objective approaches in estimating the use of new teaching approaches, such as direct observation of randomly chosen classes made by an independent trained supervisor.
2. Direct observation of teachers' lecturing can provide insight both into the quality and quantity of their employment of new techniques. Another approach for assessing quality might be administering a questionnaire to students, asking them to rank teachers according to their teaching skills. If those teachers who never attended the training are consistently ranked lower that the trained, the result might indicate different quality of lecturing.

Reliability

Findings:

- The measurement has been made only once. The report is focused on findings, but only few words are written on details about the sampling, procedure, rate of refusal and data cleaning.

Recommendations for improvement:

1. More detailed description of the research procedure in the report might ensure more stable and comparable replications of data collection over time.

Timeliness

Findings:

- The school year starts in September and the survey was administered one month later. Neither students have enough time to make judgments, nor teachers are given a chance to implement the methods sufficiently.

Recommendations for improvement:

1. Surveys should be done in November.

Practicality

Findings:

- This indicator, as it is measured now, is most of all practical as data can be obtained frequently and for reasonable cost. Its practicality though is at the cost of its directness (see validity above).

ii. *Increased number of teachers using improved methodology as per the EDC activity*

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Objectivity:

Findings:
- The definition is vague. The PMP does not say how many lessons per particular period have to be delivered by using the new methodology so that a teacher will belong to the group who use it. The interviewed implementer was not sure either.
- If the measurement unit is "number of lecture plans prepared during a certain period of time", there is no ambiguity over what kind of data should be collected.
- As far as the reports from the Change Groups are concerned, the methodology of data collection is not yet developed.

Recommendations for improvement:
1. The precise definition of the indicator should be improved so that operationally it is clear at what point a teacher can be counted as using the new methodology.

b. Increased number of career centers

There are no significant findings or recommendations regarding this indicator.

3. IR 3.4.2: Support a training system for professionalization of school directors

a. Increased number of school directors complete certification training

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:
- The target is defined as 100% of directors. Its completion can be threatened by two things: (a) that some directors are not motivated to accomplish the training (e.g. those close to the end of their mandate or pension) and (b) that another teacher certification training may become available from another training organization.

Recommendations for improvement:
1. The SO Team should modify this definition by changing the indicator's targets to allow for these possible factors.
2. The DQA Tam is recommending this indicator be included in the Annual Report, modified as described above.

4. IR 3.4.3: Increased access to education

a. Number of students enrolled in Southeastern Europe University

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Validity

Findings:
- The indicator is only measuring access to higher education.
- There are no SO 3.4 activities addressing higher education or designed to increase student enrollment in SEEU.

Recommendations for improvement:
1. The SO Team should consider re-wording the IR to limit it to access to higher education.
2. The SO Team should either design and introduce activities directed at higher education and intended to increase enrollment in SEEU or it should drop this indicator and IR.

5. IR 3.4.4: Local economic development

a. Sustained business growth at local level

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each. As is noted under "definition" the indicator is, in fact, two distinct indicators. Recommendations are made regarding how to deal with the multi-dimensionality of the indicator as is currently framed. The DQA Team did, however, do an assessment of one of these two parts. The findings and recommendations that follow "definition" are in relation to what is shown in the PMP for SO 3.4 as IR 3.4.4.a (a) which is focused on *"value of increased sales of project assisted association members in the sheep and fruits and vegetables sector (sic)."*

Definition

Findings:
- This indicator is really two indicators with two different IPs, Activity Managers, objects of focus, definitions and targets. There is no attempt to aggregate them.
- **Regarding *"value of sales…in the sheep and fruits/vegetable sectors"*** the method used by the IP for calculating sales, which is quite clear and precise in reality, is not given in its Performance Monitoring and Evaluation Plan, even though it does note as a "data limitation" that "private companies in Macedonia tend to be secretive regarding their financial performance".

Recommendations for improvement:
1. Break the indicator into two distinct indicators.
2. **Regarding *"value of sales…in the sheep and fruits/vegetable sectors"*** the IP should provide a clear description. How has this noted limitation been dealt with?

Objectivity

Findings:
- Different members of the SO team were unclear about how to interpret the sales reported here as compared with the "sales" reported by another IP for such clusters as cheese, IT and wine. (And it was unclear during the DQA if they also include export sales.) This is not so much a findings about the objectivity (clarity) of the indicator as about the SO team members' varied interpretation of it.

Recommendations for improvement:
1. The SO Team Leader should make clear to SO team members which sales data are domestic and which are export.

b. Internet users per 10,000 people

No data quality assessment was done on this indicator.

c. SMEs use of ICT to improve their competitiveness

Only the following data quality assessment criteria apply to this indicator. Findings and recommendations are given for each.

Decision

Findings:
- The indicator is framed as a result. The PMP does specify in its definition that it is to be measured as *"Number of E-Biz Centers opened and operational."*
- The project has fairly clear and concise definition of what it means to be "operational" but it is not included in the PMP.

Recommendations for improvement:
1. Change the working of the indicator to be as it actually is: *"Number of E-Biz Centers opened and operational."*
2. Include the definition of "operational" in the PMP.

Validity

Findings:
- The indicator is a better measure of activity level results than of the intermediate result it is intended to measure. Although it is direct, it is very distant from "local economic development"

Recommendations for improvement:
1. The SO team should consider another indicator for this one. In conversations with the IP, it was learned that a higher level expectation for the activity is something on the order of *"increased number of jobs created due to the E-Biz Centers"* or even more to the point of the activity *"number of jobs saved"* although certainly more difficult to measure. But something like either of these would be a better measure at the IR level.

Adequacy

Findings:
- It is not an effective measure of "on-trackedness" for the result for the reason that the result is at a higher level than it is measuring.
- Taken as a group the indicators currently tell a story about:
  - Increased sales in selected, somewhat small and definitely rural, sectors;
  - Number of people who use the Internet per 10,000; and
  - Increase in the number of E-Biz Centers.

(See each of the Indicator Checklists.)
Together these are not a very comprehensive measure of "local economic development."

Recommendations for improvement:
1. The SO Team should consider defining a more comprehensive measure of employment created by E-Biz Centers to raise the adequacy of these indicators as a package.

III RECOMMENDAIONS FOR ANNUAL REPORT INDICATORS

A. Introduction

The SO 1.3 Team has pretty much selected the indicators they want to include in the Annual Report. The SO 2.0 Team has identified certain indicators that they think would be better than others. The SO 3.4 Team has not selected or identified

preferred indicators to be included in the Annual Report although there probably is consensus around at least one in the education area. Since the DQA Team has considered all of the indicators, it makes sense (and indeed the team was tasked) to make recommendations for those indicators most appropriate for including in the Annual Report. Criteria used by the DQA Team Leader included:

- The preferences of the SO teams;
- The quality of the indicators;
- The comprehensiveness of the indicator and
- The level of measurement of the indicators.

In making the final selection, the Mission – overall – needs to reach agreement on the location of the set of indicators currently under the Intermediate Result labeled "local economic growth". This IR is sometimes seen as within SO 1.3 and sometimes seen as belonging to SO 3.4. As far as the Annual Report is concerned, the issue is: if there is an indicator that reflects "local economic growth" to be included in the AR, then under which SO should it be reported?

B. SO 1.3: Accelerated Development and Growth of the Private Sector

Result	Indicator	Reason
SO 1.3: Accelerated Development and Growth of the Private Sector	(a) Total full time employment in private sector firms	Selected by SO Team
	(b) Private sector employment as a percentage of total employment	Recommended by DQA Team
IR 1.3.1: Bank and non-bank financial institutions strengthened	(a) Bank lending relative to bank capacity to lend	
	(b) Securities markets and non-bank financial institutions	Selected by SO Team
	(c) Market turnover	
IR 1.3.2: Private sector firms more competitive	(a) Total exports	
	(b) Foreign direct investment	
	(c) Total exports of sectors assisted by USAID	Selected by SO Team
IR 1.3.3: Enabling environment for investment improved	(a) Speed of business registration	
	(b) Shareholder awareness of their rights	
	(c) Additional taxpayers	Recommended by DQA Team
	(d) Compliance to WTO requirements	

C. SO 2.0: More Legitimate Democratic Institutions

Result	Indicator	Reason
SO 2.0: More Legitimate Democratic Institutions	(a) Public perception of respect for democratic values	Preferred by SO Team
	(b) Public perception of effectiveness of key government institutions	Preferred by SO Team
IR 2.0 1: Increased citizen participation in political and social decision making	(a) Public participation in political activities	Recommended by DQA Team
	(b) Continuing initiatives at the local level	Preferred by SO Team
	(c) Women's participation in community level decision making	
	(d) NGO Sustainability Index	Preferred by SO Team
IR 2.0 1.1: Citizen attitudes towards democratic processes and practices improved	(a) Public trust in NGOs	
IR 2.0.1.2: Improved opportunities for citizen participation	(a) Public participation in civil society organizations	
IR 2.0.1.3: Improved citizen access to objective, unbiased information	No Indicator	
IR 2.0.2.1: More effective judiciary	(a) Public perception of the effectiveness of courts, both nationally an in the pilot court areas	
	(b) Freedom House "Rule of Law" rating	
IR 2.0.2.1: More effective judiciary	(a) Civil case backlog reduction in pilot courts	Preferred by SO Team
	(b) User satisfaction with courts, in pilot courts	
	(c) Changes in the performance of the justice system	
IR 2.0.2.2: More effective legal personnel	(a) User satisfaction with judges, prosecutors and attorneys in pilot courts	
	(b) Changes in the performance and ethics of non-judicial professions	
IR 2.0 2.3: Laws, regulations and policies that support market-based economy adopted	(a) Public trust in local government	
IR 2.0.3: More effective, responsive and accountable local government	(a) Key laws and implementing regulations adopted	
IR 2.0.3.1: Local governments have increased responsibility and financial resources	No Indicator	
IR 2.0.3.2: Effective relationship between citizens and local government exits	(a) Public participation in local government decision making	

Result	Indicator	Reason
IR 2.0.3.3: Municipal associations satisfactorily serve the interests of their members	No Indicator	
IR 2.0.3.4: Local governments improve the management of municipal services	No Indicator	
IR 2.0.4: Increased confidence in government institutions and political processes	(a) Trust in Parliament	
IR 2.0.4.1: Parliament functions in a more effective and visibly representative manner	No indicator	
IR 2.0.4.2: Degree to which elections are free and fair improved	(a) Free and fair elections held	
IR 2.0.4.3: Political party transformed to a platform and performance basis	No Indicator	
IR 2.0.4.4: Communication between elements of government and citizens improved	No Indicator	

D. SO 3.4: Mitigate Social Impact of the Transition to Market-Based Democracies

Result	Indicator	Reason
SO 3.4: Mitigate Adverse Social Impact of the Transition to Market-Based Democracies	Indicators for "Youth better prepared for employment through education programs" being developed: • Instructional practice; • Career preparation; • Computer skills; • Computer access; and • Problem solving skills.	Recommended by DQA Team
IR 3.4.1: Improved quality and relevance of instruction at primary and secondary schools	(a) Increased number of teachers using improved methodology	
	(b) Increased number of career centers	
IR 3.4.2: Support a training system for professionalization of school directors	(a) Increased number of school directors complete certification training	Recommended by DQA Team
IR 3.4.3: Increased access to education	(a) Number of students enrolled in Southeastern Europe University	
IR 3.4.4: Local economic development	(a) Sustained business growth at local level: Assisted members in sheep, fruits and vegetables and assisted artisan members	
	(b) Internet users per 10,000 people	
	(c) SMEs use of ICT to improve their competitiveness	

IV TRAINING OF MACEDONIAN DQA SPECIALISTS

A. Introduction

The Mission requested that two Macedonia research specialist be trained in DQA methods by the Team Leader. The thought was that university professors or research association professionals skilled in socio-economic research methods would be good candidates for the DQA. Once trained, they would be a resource for the Mission.

As it happened, the timing of the DQA did not fit well with the availability of these types of people. Final university exams fall in June and all the candidates were giving and grading exams. However, while none of the candidates were available during the full 18 days of the DQA three candidates were available for most of that period. The Mission decided to use three Macedonian research specialists - Nikolina Kenig, Ilija Todorovski and Anica Dragovic - on a part-time basis. (See Attachment D for their resumes.)

The Team Leader took one and a half days early in the DQA to train the Macedonian DQA Team members, mostly in a workshop environment, in the following topics:

(1) USAID performance monitoring methods including
- Results frameworks,
- Performance Management Plans (PMP),
- Strategic objectives and intermediate results and
- Performance indicators;

(2) Data quality criteria for performance indicators including
- Validity,
- Reliability,
- Timeliness,
- Practicality and
- Adequacy and

(3) USAID reporting schedules and requirements.

The Macedonian DQA Team members then chose from amongst the thirty eight subject indicators of the DQA those that most fit their interests and technical background. They studied the PMPs and relevant background documents for their indicators and participated in some of the USAID and implementing partner staff interviews with the Team Leader. Finally, they completed DQA checklists for the indicators they were covering. The Team Leader was able to include much of their findings in the final checklists included in Attachment C.

On Macedonian Team member was unable to finish the DQA due to an emergency in her family.

B. Strengths of the Macedonia DQA Team Members

1. The level of skills in behavioral research and socio-economic issues such as democracy building in a transitional situation and education and economic growth that the Macedonian DQA Team members brought with them was very high. These basic research skills include
   - Survey research methods,
   - Statistical analysis methods,
   - Data collection methods, and
   - Sampling
2. The level of their understanding of and familiarity with secondary sources of data available in Macedonia was very sharp and current. They knew most secondary data sources used in the three PMPs and understood their strengths and weaknesses.
3. They have worked enough with other donors to be quite familiar with the general performance monitoring methodologies that their adaptation to and understanding of USAID's measurement of results through indicators was very quick.
4. They were professionally familiar with the principal research contractors that USAID uses for its Citizen Attitude Surveys and other research efforts which made their access to those firms quick and easy.
5. The Macedonian DQA Team members brought a neutral perspective to the indicators they focused on. They could look at the indicators with a greater freshness than the SO Team members. As they work with the Mission in the future this neutrality and freshness will complement SO team perspectives which may be more limited by the history of the indicators.

C. Weaknesses in the Training

1. There is a lot about USAID's overall management system that gives the DQA a context within which the DQA "makes sense". The Macedonian Team members lacked this organizational management context. The principal management features peculiar to USAID and which the team lacked included such things as

   - Measuring the achievement of organizational results through indicators,
   - The hierarchy of results in USAID's system,
   - Strategic objective teams and their role,
   - The strategic planning process in which results are determined,
   - The reporting relationship between missions and USAID/Washington and who answers to whom and
   - The relationship of mission staff with implementing partner staff and the roles each play in achieving results.

   It was not possible during the training to give the Macedonian DQA Team members a clear enough understanding of this organizational analytical context.

   If there is a chance to send the members to training workshops in PMP, strategic planning, evaluation or activity management it would be helpful.

Also, the more they participate in performance monitoring activities for the Mission the greater understanding they will acquire.

2. This one training exercise will not be enough to enable the Macedonian DQA Team members to retain the knowledge they have gained. The Mission should use them as soon as possible for related performance monitoring efforts. An opportunity may present itself in the Fall when a new municipal government activity will begin. The Macedonian team could help the SO 2.0 Team develop new IR 2.0.3 indicators.

ATTACHMENT A

Scope of Work

Persons Interviewed

PERSONS INTERVIEWED

USAID Officials

Thomas Mehen, Program Development Advisor
Ivica Vasev, Program Development Specialist

Democracy and Local Governance Office
Michael Eddy, General Development Officer
Afrodita Salja, Project Management Specialist
Melita Cokrevska, Project Management Specialist
Antoaneta Skartova, Project Management Specialist

Private Sector Office
Geoffrey Minott, Private Sector Officer
Jovan Madjovski, Project Management Specialist
Zdravko Sami, Project Management Specialist
Meri Cuculovska, Project Management Specialist

Social Transition Office
Cecilia Sun, Education Officer
Elizabeth Markovic, Workforce Project Specialist
Lela J Jakovleska, Education Specialist
Natasha Murdzeva, Education Program Assistant

Implementing Partners

SO 1.3: Accelerated Development and Growth of the Private Sector

Lester H. Sweeting, ESQ	Fiscal Reform Project, Chief of Party
	Macedonia Corporate Governance/Company Law Proj.,
Dr. Gregory F. Maassen	Chief of Party
Snezana ??????	Research Manager
Thomas Carson, Ph.D.	Research Director
	Macedonia Competitiveness Activity
Suzi Kanya Hagan	Chief of Party
Kevin Newman	Competitiveness Advisor
Anthony ??????	Sheep and Cheese Cluster Advisor
Dane Smith	Research Director
	Macedonia Agribusiness Marketing Activity,
David Blood	Chief of Party
Kristina Deriban	Monitoring and Evaluation Manager
	WTO Compliance Activity
Zoran Jolevski, Ph.D.	National WTO Coordinator
Gordana Toseva	Senior Legal Advisor

SO 2.0: More Legitimate Democratic Institutions

	Parliamentary Development
Chris Henshaw	Chief of Party
Elvis ?????	Constituency Relations Chief
	Rule of Law Project
Marilyn Zelin	Country Director
Keti Ilievska	Legal Advisor
Amanda Ashford	Legal Professional Index Assessor
	Macedonia Court Modernization Project,
Douglas E. Myers	Chief of Party
Gordana Stojanova	Pilot Court Manager
	Macedonian Stock Exchange Project,
Ivan Steriev	Director, Trading Department
Margica Miova	Legal Advisor, Macedonian Financial Sector Project

SO 2.0: More Legitimate Democratic Institutions (cont.)

	Macedonia Community Self-Help Initiative,
Diannna Wuagneux Ph.D	Chief of Party
Nebojsa Mojsoski	Monitoring and Evaluation Coordinator
Natasa Stankovic	Regional Representative
	Democracy Network
Irena Stevcevska	Community Coordinator
	BRIMA (Gallup) International
Gjorgji Kimov	Research Director
Gordana ???????	???????

SO 3.4: Mitigating the Adverse Social Impacts of the Transition to the Market-Based Democracy

Spomenka Lazarevska	Creative Teaching and Learning, Project Director
Natasha ??????	??????
	Secondary Education Activity,
Mark Shapiro, Ph.D.	Research Scientist
Zoran Stojanov	Research, Monitoring and Evaluation Coordinator
	E-Schools In Macedonia Activity,
Luis Rodriguez	Chief of Party
	Macedonian ICT and Local Economic Activity Development,
	EDC, Center for Innovative Technologies,
Janice Brodman	Director
George Peterson	Chief of Party

Activity Field Staff and Clients

Meriton Abazi, Caseificio Cezarina 2 – Fejzi (Cheese production and export)

Sheep farm for cheese production

??????, National Association of Dairy Producers

Government Statistical Office, regarding the
Annual Labor Force Survey

Municipality of Staro Nagorichane/Niculjane,
Mayor Vlasta Dimkovic

Public Communal Enterprise for Municipality of Staro Nagorichane/Niculjane,

ATTACHMENT C

Data Quality Assessment Checklists:
SO 1.3;
SO 2.0 and
SO 3.4

QUALITY ASSESSMENT CHECKLIST

Intermediate Result: NA

Performance indicator: Total full time employment in private sector firms

Data source(s): Labor Force Survey and Government Statistical Office

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID:

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

1. VALIDITY—Do the data adequately represent performance?			
Face Validity		**No**	**Comments**
➢ between the activity or program and what is being measured?	**X**		Defined as "number of persons employed, in thousands, disaggregated by gender and ethnicity", this is a valid and direct measure of the Strategic Objective. The process is quite rigorous and consistently applied each year.
➢ Other			
Measurement Error			
(only applies when the data source is a survey)			Source: Labor Force Survey, State Statistical Office, Skopje
➢ Were samples representative?	**X**		The sample comprised 7200 households up to the year 2003
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?	**X**		
➢ If the instrument was self-reporting were adequate instructions provided?			Not applicable
➢ Were response rates sufficiently large?	**X**		
➢ Has non-response rate been followed up?	**X**		
Non Sampling Error			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
➢ Is the data collection instrument well designed?	**X**		
➢ Were there incentives for respondents to give incomplete or untruthful information?		**X**	
➢ Are definitions for data to be collected operationally precise?	**X**		
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?	**X**		
➢ Were there efforts to reduce the potential for personal bias by enumerators?			Not observed
1. Other			
Transcription Error			
➢ What is the data transcription process? Is there potential for error?			Not observed
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)	**X**		According to interview respondents but not observed
➢ Have data errors been tracked to their original source and mistakes corrected?	**X**		According to interview respondents but not observed
➢ If raw data need to be manipulated to produce the data required for the indicator:	**X**		
➢ Are the correct formulae being applied?	**X**		
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?	**X**		
➢ Have procedures for dealing with missing data been correctly applied?	**X**		They are not included in the calculations
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)	**X**		
1. Other			
Representativeness of Data			
➢ Is the sample from which the data are drawn representative of the population served by the activity?	**X**		
➢ Did all units of the population have an equal chance of being selected for the sample?	**X**		
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually	**X**		

1. VALIDITY—Do the data adequately represent performance?

	Yes	No	Comments
exclusive (for geographic frames)			
➢ Is the sample of adequate size?	X		
➢ Are the data complete? (i.e., have all data points been recorded?)	X		

Findings:

➢ Defined as "number of persons employed, in thousands, disaggregated by gender and ethnicity", this is a valid and direct measure of the Strategic Objective. The process is quite rigorous and consistently applied each year. The Government Statistics Office publishes two calculations for this variable: One as reported by the Government Pensions Office which is generally regarded as under-reporting employment and this as taken through an annual labor force survey.

Recommendations for improvement:

1. As important an indicator as this is for SO 1.3, the SO team should continue to monitor and understand the procedures used in the Labor Force Survey possibly by directly observing data gathering, transmission and calculation.

2. RELIABILITY—Are data collection processes stable and consistent over time?

	Yes	No	Comments
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?	X		Although not observed, the procedures are rigorously followed each year.
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?	X		
➢ Is the same sampling method used from year to year, location to location, data source to data source?	X		
2. Other			
Internal quality control			
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?	X		
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?	X		
➢ Do these procedures provide for periodic sampling and quality assessment of data?	X		
3. Other			
Transparency			
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?	X		Although not directly observed
➢ Are data problems at each level reported to the next level?	X		Respondents report a rigorous process of checking and control between levels.
➢ Are data quality problems clearly described in final reports?	X		In footnotes

Comment of the analyst:

The instruments are applied and data obtained by the State Statistical Office and analyzed. All information dealing with data gathering, field work, quality of control for this DQA were obtained through interviews with the official persons in charge for the Labor Force Survey employed at the State Statistical Office. In the introduction of the Survey it is said that the Labor Force Survey is conducted in accordance to the methodological recommendation of the International Labor Organization (ILO) ratified on the 13 International Conference of Labor Statistician and the recommendation of the European Statistical Bureau (Eurostat). We would like to convey the statement of the interviewees that the Survey is designed and implemented according to the highest international standards.

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		Until the year 2003 the data were collected once a year throughout the whole territory of the country and included approximately 1.5% of the total number of households. The results were published a month and a half after the data gathering. Since the beginning of the year 2004 there is an attempt to gather data more frequently through four surveys taken during the year, on quarterly basis. The coverage of each survey is 5,000 households, that means all together in a course of a year 20,000 households will be surveyed. The first and third surveys cover the identical population as do the second and the fourth surveys. Beginning in 2005 the results will be published once a year, a month and a half after the fourth survey.
➢ Is a regularized schedule of data collection in place to meet program management needs?	**X**		
2. Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	**X**		Not applicable
3. Other			
Currency			
➢ Are the data reported in a given timeframe the most current practically available?	**X**		The data are published a month and a half after their gathering.
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	**X**		
➢ Are the data reported as soon as possible after collection?	**X**		
➢ Is the date of collection clearly identified in the report?	**X**		

3. TIMELINESS—Are data collected frequently and are they current?

	Yes	No	Comments

Findings:

- Until the year 2003 the data were collected once a year throughout the whole territory of the country and included approximately 1.5% of the total number of households. The results were published a month and a half after the data gathering. Since the beginning of the year 2004 there is an attempt to gather data more frequently through four surveys taken during the year, on quarterly basis. The coverage of each survey is 5,000 households, that means all together in a course of a year 20,000 households will be surveyed. The first and third surveys cover the identical population as do the second and the fourth surveys. Beginning in 2005 the results will be published once a year, a month and a half after the fourth survey.

Recommendations for improvement:

1. It would be more convenient for policy makers and other stakeholders if the Labor Force Survey were published on semi-annual basis where the results covering the first and second survey will be preliminary ones and those of the fourth and final surveythe definite ones.

4. OBJECTIVITY — Can it be independently recognized, at "face value", as an acceptable measure of the desired result?

	Yes	No	Comments
➢ Is it unambiguous about what is being measured?	X		
➢ Is there general agreement over the interpretation of the results?	X		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	X		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	X		
➢ Other			

Recommendations for improvement:

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?

	Yes	No	Comments
➢ Are timely data available (i.e., is data current and available on regular basis)?	X		
➢ Can the data be collected frequently	X		See **Timeliness** above

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	**No**	**Comments**
enough to inform management decisions?			
➢ Are the costs of data collection reasonable?	**X**		
➢ Are data gathered from secondary, e.g. government, sources?	**X**		
➢ If so, are data reliable, particularly in the way they are gathered and calculated?	**X**		
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?	**X**		Please, see **Validity.**
➢ Other			
Recommendations for improvement:			

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?

	Yes	No	Comments
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?		X	
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	X		This indicator should be seen together with SO 1.3(b) and not separately. In order to best interpret what is going on with employment in the private sector both the absolute and relative measures of it are necessary. For the SO 1.3 team it would also be interesting (if not necessary) to view these along with total unemployment.
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		X	
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?		X	
➢ Other aspect of "sufficiency?		X	
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?		X	
➢ Impact of services	X		
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			

Recommendations for improvement:

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: NA

Performance indicator: Private sector employment as a percentage of total employment

Data source(s): Labor Force Survey and Government Statistical Office

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report? RECOMMENDED

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID:

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

1. VALIDITY—Do the data adequately represent performance?			
Face Validity	**Yes**	**No**	**Comments**
➢ Is there a solid, logical relation between the activity or program and what is being measured?	**X**		
➢ Other			
Measurement Error			
Sampling Error (only applies when the data source is a survey)			Source: Labor Force Survey, State Statistical Office, Skopje
➢ Were samples representative?	**X**		The sample comprised 7200 households up to the year 2003
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?	**X**		
➢ If the instrument was self-reporting were adequate instructions provided?			Not applicable
➢ Were response rates sufficiently large?	**X**		
➢ Has non-response rate been followed up?	**X**		
Non Sampling Error			
➢ Is the data collection instrument well designed?	**X**		
➢ Were there incentives for respondents		**X**	

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?	**X**		
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?	**X**		
➢ Were there efforts to reduce the potential for personal bias by enumerators?			Not observed
1. Other			
Transcription Error			
➢ What is the data transcription process? Is there potential for error?			Not observed
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)	**X**		According to interview respondents but not observed
➢ Have data errors been tracked to their original source and mistakes corrected?	**X**		According to interview respondents but not observed
➢ If raw data need to be manipulated to produce the data required for the indicator:	**X**		
➢ Are the correct formulae being applied?	**X**		
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?	**X**		
➢ Have procedures for dealing with missing data been correctly applied?	**X**		They are not included in the calculations
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)	**X**		
2. Other			
Representativeness of Data			
➢ Is the sample from which the data are drawn representative of the population served by the activity?	**X**		
➢ Did all units of the population have an equal chance of being selected for the sample?	**X**		
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)	**X**		
➢ Is the sample of adequate size?	**X**		
➢ Are the data complete? (i.e., have all	**X**		

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments
data points been recorded?)			

Recommendations for improvement:

1. As important an indicator as this is for SO 1.3, the SO team should continue to monitor and understand the procedures used in the Labor Force Survey possibly by directly observing data gathering, transmission and calculation.

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	No	Comments
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?	X		Although not observed, the procedures are rigorously followed each year.
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?	X		
➢ Is the same sampling method used from year to year, location to location, data source to data source?	X		
3. Other			
Internal quality control			
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?	X		
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?	X		
➢ Do these procedures provide for periodic sampling and quality assessment of data?	X		
4. Other			
Transparency			
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?	X		Although not directly observed
➢ Are data problems at each level reported to the next level?	X		Respondents report a rigorous process of checking and control between levels.
➢ Are data quality problems clearly described in final reports?	X		In footnotes

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	No	Comments
Comment of the analyst: The instruments are applied and data obtained by the State Statistical Office and analyzed. All information dealing with data gathering, field work, quality of control for this DQA we re obtained through interviews with the official persons in charge for the Labor Force Survey employed at the State Statistical Office. In the introduction of the Survey it is said that the Labor Force Survey is conducted in accordance to the methodological recommendation of the International Labor Organization (ILO) ratified on the 13 International Conference of Labor Statistician and the recommendation of the European Statistical Bureau (Eurostat). We would like to convey the statement of the interviewees that the Survey is designed and implemented according to the highest international standards.			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		Until the year 2003 the data were collected once a year throughout the whole territory of the country and included approximately 1.5% of the total number of households. The results were published a month and a half after the data gathering. Since the beginning of the year 2004 there is an attempt to gather data more frequently through four surveys taken during the year, on quarterly basis. The coverage of each survey is 5,000 households, that means all together in a course of a year 20,000 households will be surveyed. The first and third surveys cover the identical population as do the second and the fourth surveys. Beginning in 2005 the results will be published once a year, a month and a half after the fourth survey.
➢ Is a regularized schedule of data collection in place to meet program management needs?	**X**		
➢ Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	**X**		Not applicable
5. Other			
Currency			
➢ Are the data reported in a given timeframe the most current practically available?	**X**		The data are published a month and a half after their gathering.
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	**X**		
➢ Are the data reported as soon as possible after collection?	**X**		
➢ Is the date of collection clearly identified in the report?	**X**		

3. TIMELINESS—Are data collected frequently and are they current?

	Yes	No	Comments

Findings:

- Until the year 2003 the data were collected once a year throughout the whole territory of the country and included approximately 1.5% of the total number of households. The results were published a month and a half after the data gathering. Since the beginning of the year 2004 there is an attempt to gather data more frequently through four surveys taken during the year, on quarterly basis. The coverage of each survey is 5,000 households, that means all together in a course of a year 20,000 households will be surveyed. The first and third surveys cover the identical population as do the second and the fourth surveys. Beginning in 2005 the results will be published once a year, a month and a half after the fourth survey.

Recommendations for improvement:

1. It would be more convenient for policy makers and other stakeholders if the Labor Force Survey were published on semi-annual basis where the results covering the first and second survey will be preliminary ones and those of the fourth and final survey the definite ones.

4. OBJECTIVITY — Can it be independently recognized, at "face value", as an acceptable measure of the desired result?

	Yes	No	Comments
➢ Is it unambiguous about what is being measured?	X		
➢ Is there general agreement over the interpretation of the results?	X		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	X		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	X		
➢ Other			

Recommendations for improvement:

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	**No**	**Comments**
➢ Are timely data available (i.e., is data current and available on regular basis)?	**X**		
➢ Can the data be collected frequently enough to inform management decisions?	**X**		See **Timeleness**
➢ Are the costs of data collection reasonable?	**X**		
➢ Are data gathered from secondary, e.g. government, sources?	**X**		
➢ If so, are data reliable, particularly in the way they are gathered and calculated?	**X**		
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?	**X**		Please, see **Validity**
➢ Other			
Recommendations for improvement:			

6 ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?		**X**	
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	**X**		This indicator should be seen together with SO 1.3 (a) and not separately. In order to best interpret what is going on with employment in the private sector both the absolute and relative measures of it are necessary. For the SO 1.3 team it would also be interesting (if not necessary) to view these along with total unemployment.
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		**X**	
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?		**X**	
➢ Other aspect of "sufficiency?		**X**	

6 ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?

	Yes	No	Comments
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?			
➢ Impact of services	X		
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			

Findings:

➢ Although not relevant to the quality of this indicator, the SO 1.3 team has not selected this indicator for inclusion in the Annual Report while its counterpoint indicator SO 1.3 (a) was selected.

Recommendations for improvement:

1. This indicator should be seen together with SO 1.3 (a) and not separately. In order to best interpret what is going on with employment in the private sector both the absolute and relative measures of it are necessary. For the SO 1.3 team it would also be interesting (if not important) to view both these indicators along with total unemployment. He DQA Team is recommending that this indicator be included in the Annual Report and presented together with SO 1.3 (a).

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Bank and non-bank financial institutions strengthened

Performance indicator: Bank lending relative to bank capacity to lend

Data source(s): Balance Sheet from National Bank of the Republic of Macedonia (NBRM)

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID officials participating or consulted: Jovan Madjovski and Zdravko Sami

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

1. VALIDITY—Do the data adequately represent performance?

	Yes	No	Comments
Face Validity			
➢ Is there a solid, logical relation between the activity or program and what is being measured?	X		
Directness			
➢ Does it closely measure the result it is intended to measure?	X		
➢ Is it grounded in theory and practice?	X		This is a generally accepted measure of strengthened capacity of the banking sector.
➢ Does it represent an acceptable measure to both proponents and skeptics?	X		
Measurement Error			Not applicable
Sampling Error (only applies when the data source is a survey)			
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
up?			
Non Sampling Error			
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			Not applicable
➢ What is the data transcription process? Is there potential for error?			
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with missing data been correctly applied?			
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments
units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			
Recommendations for improvement:			

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	No	Comments
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?			The definition of deposits and loans and, precisely, what are included in each changed in 2001 but the baseline from 1999 and the actuals from 2000 were adjusted in the PMP.
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?	X		NBRM Balance Sheet
➢ Is the same sampling method used from year to year, location to location, data source to data source?			Not applicable
➢ Other			
Internal quality control			Not applicable
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			
➢ Do these procedures provide for periodic sampling and quality assessment of data?			
➢ Other			
Transparency			Not applicable
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?			
➢ Are data problems at each level reported to the next level?			
➢ Are data quality problems clearly described in final reports?			
Recommendations for improvement:			

3. TIMELINESS—Are data collected frequently and are they current?

	Yes	No	Comments
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	X		
➢ Is a regularized schedule of data collection in place to meet program management needs?			Not applicable
➢ Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	X		
➢ Other			
Currency	X		
➢ Are the data reported in a given timeframe the most current practically available?			
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	X		
➢ Are the data reported as soon as possible after collection?			Not applicable
➢ Is the date of collection clearly identified in the report?	X		

Recommendations for improvement:

4. OBJECTIVITY — Can it be independently recognized, at "face value", as an acceptable measure of the desired result?

	Yes	No	Comments
➢ Is it unambiguous about what is being measured?	X		
➢ Is there general agreement over the interpretation of the results?	X		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	X		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	X		
➢ Other			

Recommendations for improvement:

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	**No**	**Comments**
➢ Are timely data available (i.e., is data current and available on regular basis)?	**X**		
➢ Can the data be collected frequently enough to inform management decisions?	**X**		
➢ Are the costs of data collection reasonable?	**X**		
➢ Are data gathered from secondary, e.g. government, sources?	**X**		
➢ If so, are data reliable, particularly in the way they are gathered and calculated?	**X**		
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?	**X**		
➢ Other			
Recommendations for improvement:			

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?		**X**	Insofar as all the activities under IR1.3.1 are aimed at strengthening bank and non-bank institutions, this is a measure of the ultimate purpose of the result.
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	**X**		There are three indicators measuring this IR this indicator covers the banking sector, IR1.3.1(b) addresses the non-banking sector while IR1.3.1 (c) looks at the major financial market institution in Macedonia – the stock exchange.
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		**X**	
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?		**X**	
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome of the program, as opposed to the	**X**		

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
completion of an activity or process?			
➢ Impact of services	**X**		
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			
Recommendations for improvement:			

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Bank and non-bank financial institutions strengthened

Performance indicator: Securities markets and non-bank financial institutions

Data source(s):

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report? YES

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID officials participating or consulted: Jovan Madjovski and Zdravko Sami

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

1. VALIDITY—Do the data adequately represent performance?			
Face Validity	**Yes**	**No**	**Comments**
➢ Is there a solid, logical relation between the activity or program and what is being measured?	**X**		This indicator is taken from the European Bank for Reconstruction and Development (EBRD) annual report.

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
➢ Other			
Measurement Error			Not applicable
Sampling Error (only applies when the data source is a survey)			
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed up?			
Non Sampling Error			
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			Not applicable
➢ What is the data transcription process? Is there potential for error?			
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments
missing data been correctly applied?			
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			

Findings:

➢ This indicator is taken from the European Bank for Reconstruction and Development (EBRD) annual report. The DQA Team attempted to study the definitions, data collection methods and calculations used by the EBRD only to learn that all data used were themselves gathered from secondary sources, compiled and included in the report from London. Thus, no DQA, through interviews, analysis of definitions and data gathering methods, or direct observation, was done on this indicator.

Recommendations for improvement:

1. The SO Teams should find a way to learn more about how these data are collected and calculated as well as the precise definitions that are used in order to better interpret the data.

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	No	Comments
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?	X		
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?			Not observed
➢ Is the same sampling method used from year to year, location to			Not observed

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
location, data source to data source?			
➢ Other			
Internal quality control			
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			Not observed
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			Not observed
➢ Do these procedures provide for periodic sampling and quality assessment of data?			Not observed
➢ Other			
Transparency			
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?			Not applicable
➢ Are data problems at each level reported to the next level?			Not observed
➢ Are data quality problems clearly described in final reports?			Not observed
Recommendations for improvement:			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		Annually
➢ Is a regularized schedule of data collection in place to meet program management needs?			Not observed
➢ Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?			Not observed
➢ Other			
Currency			
➢ Are the data reported in a given timeframe the most current practically available?	**X**		
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	**X**		
➢ Are the data reported as soon as	**X**		

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	No	Comments
possible after collection? ➢ Is the date of collection clearly identified in the report?	X		
Recommendations for improvement:			

4. OBJECTIVITY — Can it be independently recognized, at "face value", as an acceptable measure of the desired result?			
	Yes	No	Comments
➢ Is it unambiguous about what is being measured?	X		
➢ Is there general agreement over the interpretation of the results?	X		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	X		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	X		
➢ Other			
Recommendations for improvement:			

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	No	Comments
➢ Are timely data available (i.e., is data current and available on regular basis)?	X		
➢ Can the data be collected frequently enough to inform management decisions?	X		
➢ Are the costs of data collection reasonable?	X		
➢ Are data gathered from secondary, e.g. government, sources?	X		
➢ If so, are data reliable, particularly in the way they are gathered and calculated?			Not observed
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?			Not observed
➢ Other			

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	**No**	**Comments**
Recommendations for improvement:			

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?		**X**	Insofar as all the activities under IR1.3.1 are aimed at strengthening bank and non-bank institutions, this is a measure of the ultimate purpose of the result.
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	**X**		There are three indicators measuring this IR this indicator covers the non-banking sector, IR1.3.1 (a) addresses the banking sector while IR1.3.1 (c) looks at the major financial market institution in Macedonia – the stock exchange.
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		**X**	
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?		**X**	
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?	**X**		
➢ Impact of services	**X**		
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			
Recommendations for improvement:			

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Bank and non-bank financial institutions strengthened

Performance indicator: Market turnover

Data source(s): Stock Exchange Bulletin

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID officials participating or consulted: Jovan Madjovski

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

---

1. VALIDITY—Do the data adequately represent performance?			
Face Validity	**Yes**	**No**	**Comments**
➢ Is there a solid, logical relation between the activity or program and what is being measured?	**X**		This is a valid measure of the intended result: "Bank and non-bank financial institutions strengthened". The logic is that as the volume of transactions has increased, the Macedonian Stock Exchange has been able to process those transactions. However, it does not measure the quality with which those transactions are handled or the efficiency, perhaps a more valid measure. It is an input measure of capacity. The indicator measures the volume of transactions which is more validly a measure of the health of the economy and financial markets. The fact that it has gone up in the last three or for years does not necessarily mean that the capacity of the Macedonia Stock Exchange has commensurately increased.
Usefulness:			
➢ Can changes in the value of the indicator be reasonably attributed to the efforts of USAID and its partners?		**X**	The changes in the indicator from one year to another have nothing to do with USAID interventions.
➢ Does the indicator effectively tell the mission if it is "on track" regarding expected performance to -date in the implementation of the program plan?	**X**		Indirectly: That the Macedonian Stock Exchange has not broken down under the weight of greatly increased volume of transactions.
➢ Does it help the mission know about the existence possible problems in			Not observed

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
implementation?			
➢ Does it help the mission locate the sources of problems in implementation?			Not observed
➢ Is the indicator easy to understand as stated?			It's actually a bit deceptive: It seems to say something that it really isn't.
➢ Is the indicator a credible measure of the intended result?	**X**		Indirectly, as described above.
Measurement Error			Not applicable
Sampling Error (only applies when the data source is a survey)			
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed up?			
Non Sampling Error			
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			Not applicable
➢ What is the data transcription process? Is there potential for error?			
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments
applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with missing data been correctly applied?			
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			

Findings:

➢ Although the indicator is a valid measure of the result, it is an input indicator and it is not a very direct measure of the result – increased capacity of the Macedonian Stock Exchange. Moreover, although it measures the volume of transactions handled it does not measure the quality with which those transactions are handled or the efficiency, perhaps a more valid measure. The indicator measures the volume of transactions which is more validly a measure of the health of the economy and financial markets. The fact that it has gone up in the last three or for years does not necessarily mean that the capacity of the Macedonia Stock Exchange has commensurately increased.

Recommendations for improvement:

1. The SO teams should strongly consider measuring the efficiency of the Macedonian Stock Exchange, thus incorporating the costs that it requires to process transactions; a much more direct measure of the intended result.

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	No	Comments
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source	X		

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
(if data come from different sources)?			
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?			Not applicable
➢ Is the same sampling method used from year to year, location to location, data source to data source?			Not applicable
➢ Other			
Internal quality control			Not applicable
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			
➢ Do these procedures provide for periodic sampling and quality assessment of data?			
➢ Other			
Transparency			Not applicable
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?			
➢ Are data problems at each level reported to the next level?			
➢ Are data quality problems clearly described in final reports?			
Recommendations for improvement:			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		
➢ Is a regularized schedule of data collection in place to meet program management needs?			Not applicable
6. Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?			Not applicable
➢ Other			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Currency			
➢ Are the data reported in a given time frame the most current practically available?	**X**		
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)			Not applicable
➢ Are the data reported as soon as possible after collection?			Not applicable
➢ Is the date of collection clearly identified in the report?	**X**		
Recommendations for improvement:			

4. OBJECTIVITY—Can it be independently recognized, at "face value", as an acceptable measure of the desired result?			
	Yes	**No**	**Comments**
➢ Is it unambiguous about what is being measured?		**X**	The fact that the measure is going up does not mean that capacity is going up equally. (See validity above.)
➢ Is there general agreement over the interpretation of the results?			No observed
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	**X**		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?			Not applicable
➢ Other			
Recommendations for improvement:			

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	**No**	**Comments**
➢ Are timely data available (i.e., is data current and available on regular basis)?	**X**		
➢ Can the data be collected frequently enough to inform management decisions?	**X**		
➢ Are the costs of data collection reasonable?	**X**		
➢ Are data gathered from secondary,			

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	**No**	**Comments**
e.g. government, sources?			
➢ If so, are data reliable, particularly in the way they are gathered and calculated?			
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?			
➢ Other			
Recommendations for improvement:			

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?	**X**		
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	**X**		There three measures, one for banks, one for non-banks and this one measuring the institutional capacity to process transactions.
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?			
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?			
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?			
➢ Impact of services		**X**	
➢ Quality of services		**X**	
➢ Customer (end-user) satisfaction		**X**	
➢ Physical product			
➢ Major milestone			
➢ Other			
Recommendations for improvement:			

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Private sector firms more competitive

Performance indicator: Total exports

Data source(s): National Bank Bulletin

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID:

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

1. VALIDITY—Do the data adequately represent performance?			
Face Validity	**Yes**	**No**	**Comments**
➢ Is there a solid, logical relation between the activity or program and what is being measured?	**X**		
Directness:			
➢ Does it closely measure the result it is intended to measure?	**X**		However, the extent to which changes in the indicator can be attributed to USAID and its interventions is very limited. USAID activities involve only five "clusters": Cheese and sheep dairy products; lamb; ICT; tourism and fruits and vegetables. Of these there is no program intention to increase the export of fruits and vegetables. The tourism sector is not included in "total exports. The SO team has not yet figured precisely which of the hundreds of categories of ICT goods which might be exported are considered as its program's targets. Finally, the figures on lamb exports are unreliable. Only the export of cheese and sheep dairy products on a national level accurately reflect those exports over which USAID/Macedonia can claim attribution.
Measurement Error			Not applicable

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
Sampling Error (only applies when the data source is a survey)			
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed up?			
Non Sampling Error			
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			Not applicable
➢ What is the data transcription process? Is there potential for error?			
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with missing data been correctly applied?			
➢ Are final numbers reported accurate?			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
(E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			

Findings:

➢ The extent to which changes in this indicator can be attributed to USAID and its interventions is very limited. Relevant USAID activities involve only five "clusters": Cheese and sheep dairy products; lamb; ICT; tourism and fruits and vegetables. Of these there is no program intention to increase the export of fruits and vegetables. The tourism sector is not included in "total exports". The SO team has not yet figured precisely which of the hundreds of categories of ICT goods which might be exported are considered as its program's targets. Finally, the figures on lamb exports are unreliable. Interventions in the wine cluster have not yet started. Only the export of cheese and sheep dairy products on a national level accurately reflect exports over which USAID/Macedonia can claim attribution.

Recommendations for improvement:

1. An indicator closer to the interventions that SO 1.3 carries out should be identified and should be added to the array of indicators for IR 1.3.2.
2. The SO Team should consider dropping this indicator.

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
Consistency			Not applicable
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?			

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?			
➢ Is the same sampling method used from year to year, location to location, data source to data source?			
7. Other			
Internal quality control			Not applicable
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			
➢ Do these procedures provide for periodic sampling and quality assessment of data?			
➢ Other			
Transparency			Not applicable
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?			
➢ Are data problems at each level reported to the next level?			
➢ Are data quality problems clearly described in final reports?			
Recommendations for improvement:			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		
➢ Is a regularized schedule of data collection in place to meet program management needs?	**X**		
8. Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	**X**		
➢ Other			
Currency			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	No	Comments
➢ Are the data reported in a given time frame the most current practically available?	X		
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	X		
➢ Are the data reported as soon as possible after collection?	X		
➢ Is the date of collection clearly identified in the report?	X		
Recommendations for improvement:			

4 OBJECTIVITY — Can it be independently recognized, at "face value", as an acceptable measure of the desired result?			
	Yes	No	Comments
➢ Is it unambiguous about what is being measured?	X		
➢ Is there general agreement over the interpretation of the results?	X		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	X		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	X		
➢ Other			
Recommendations for improvement:			

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	No	Comments
➢ Are timely data available (i.e., is data current and available on regular basis)?	X		
➢ Can the data be collected frequently enough to inform management decisions?	X		
➢ Are the costs of data collection reasonable?	X		
➢ Are data gathered from secondary, e.g. government, sources?	X		

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?

	Yes	No	Comments
➢ If so, are data reliable, particularly in the way they are gathered and calculated?	X		The data are prepared by the Customs Administration of the Republic of Macedonia and processed and published by the State Statistical Office of the Republic of Macedonia and National Bank of the Republic of Macedonia (NBRM)
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?	X		
➢ Currency conversion			All figures in the Bulletin are presented in US$. Since the dollar fluctuates, it is very difficult to perceive the real trends. Therefore it would be better if the figures were presented in adjusted US$ that are equivalent to values of the export expressed in EUROS.

Findings:

➢ All figures in the Bulletin are presented in US$. Since the dollar fluctuates, it is very difficult to perceive the real trends.

Recommendations for improvement:

1. The figures should be presented in adjusted US$ that are equivalent to values of the export expressed in EUROS.

6 ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?

	Yes	No	Comments
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?		X	
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	X		
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		+	
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?		X	
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome			

6 ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
of the program, as opposed to the completion of an activity or process?			
➢ Impact of services	**X**		
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			
Recommendations for improvement:			

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Private sector firms more competitive

Performance indicator: Foreign direct investment

Data source(s): National Bank Bulletin and Bulletin of Balance of Payments

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID:

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

1. VALIDITY—Do the data adequately represent performance?			
Face Validity	**Yes**	**No**	**Comments**
➢ Is there a solid, logical relation between the activity or program and what is being measured?	**X**		

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
➢ Other			
Measurement Error			Not applicable
Sampling Error (only applies when the data source is a survey)			
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed up?			
Non Sampling Error			
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			Not applicable
➢ What is the data transcription process? Is there potential for error?			
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments
missing data been correctly applied?			
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			
Recommendations for improvement:			

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	No	Comments
Consistency			Not applicable
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?			
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?			
➢ Is the same sampling method used from year to year, location to location, data source to data source?			
➢ Other			
Internal quality control			Not applicable
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			
➢ Do these procedures provide for periodic sampling and quality			

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
assessment of data?			
➢ Other			
Transparency			Not applicable
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?			
➢ Are data problems at each level reported to the next level?			
➢ Are data quality problems clearly described in final reports?			
Recommendations for improvement:			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		
➢ Is a regularized schedule of data collection in place to meet program management needs?	**X**		
9. Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	**X**		
➢ Other	**X**		
Currency	**X**		
➢ Are the data reported in a given timeframe the most current practically available?	**X**		
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	**X**		
➢ Are the data reported as soon as possible after collection?	**X**		
➢ Is the date of collection clearly identified in the report?	**X**		
Recommendations for improvement:			

4. OBJECTIVITY — Can it be independently recognized, at "face value", as an acceptable measure of the desired result?

	Yes	No	Comments
➢ Is it unambiguous about what is being measured?	X		
➢ Is there general agreement over the interpretation of the results?	X		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	X		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	X		
➢ Other			

Recommendations for improvement:

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?

	Yes	No	Comments
➢ Are timely data available (i.e., is data current and available on regular basis)?	X		
➢ Can the data be collected frequently enough to inform management decisions?	X		
➢ Are the costs of data collection reasonable?	X		
➢ Are data gathered from secondary, e.g. government, sources?	X		
➢ If so, are data reliable, particularly in the way they are gathered and calculated?	X		The data come from the Customs Administration of the Republic of Macedonia and State Statistical Office of the Republic of Macedonia
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?	X		
➢ Other			

Recommendations for improvement:

6 ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?		**X**	
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	**X**		
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		**X**	
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?		**X**	
➢ Other aspect of "sufficiency?		**X**	
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?			
➢ Impact of services	**X**		
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			
Recommendations for improvement:			

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Private sector firms more competitive

Performance indicator: Total exports of sectors assisted by USAID

Data source(s): MCA reports

Partner or contractor who provided the data (if applicable): BAH

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID officials participating or consulted: Jovan Madjovski and Zdravko Sami

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

1. VALIDITY—Do the data adequately represent performance?			
Face Validity	**Yes**	**No**	**Comments**
➢ Is there a solid, logical relation between the activity or program and what is being measured?	**X**		Measuring "exports to USAID assisted sectors" as an indicator of the IR is logical and valid. However, there is an attribution problem in that USAID activities work only with selected entities in those "assisted sectors". In the cheese sector there is identify between the sector and the USAID assisted entities in it. But this is less true for lamb exports, much less true for IT exports and less still when it comes to fruits and vegetables.
➢ Other:			The indicator is in fact several indicators. Each cluster is measured and reported separately. Therefore there is no unique measure of the indicator. There is no global measure or "total" reported.
Measurement Error			Not Applicable
Sampling Error (only applies when the data source is a survey)			
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed up?			
Non Sampling Error			
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			Not Applicable
➢ What is the data transcription process? Is there potential for error?			
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with missing data been correctly applied?			
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not Applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			

1. VALIDITY—Do the data adequately represent performance?

	Yes	No	Comments
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			

Findings:

- The validity of this indicator is diminished, in some cases significantly because of the attribution problem: USAID activities work only with selected entities in the "assisted sectors" referred to in the indicator while in most cases USAID's activities do not cover the overall sectors assisted. In the cheese sector there is identify between the sector and the USAID assisted entities in it. But this is less true for lamb exports, much less true for IT exports and less still when it comes to fruits and vegetables.
- The fruits and vegetables IP does even try to measure the sector for this indicator, rather it reports exports of assisted entities.
- The indicator is in fact several indicators. The exports of each cluster is measured and reported separately. Therefore there is no unique measure of the indicator. There is no global measure or "total" reported. Moreover, the tourism cluster is not measured in terms of exports at all.

Recommendations for improvement:

1. Regarding the coverage issue – The SO team should make a judgment about whether to report overall exports of USAID assisted sectors or exports of USAID assisted entities.
2. Regarding the multi-dimensionality of the indicator – The SO team should
3. Find a measure for the tourism sector that is comparable, even if reported separately, with the exports measure adopted for the other cluster.
4. If there is a way to aggregate the remaining cluster sector exports that should be done. If this doesn't make economic sense, then they should be simply reported separately. However, because of the attribution problem noted above, if they are reported separately there must also be some indication of the extent to which the cluster entities cover the reported sector exports.

2. RELIABILITY—Are data collection processes stable and consistent over time?

	Yes	No	Comments
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?		X	Export data reported for fruits and vegetables by Land 'o Lakes is for cluster entities only while export data reported for cheese, lamb and IT by MAC/BAH are for overall sector exports.
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?		X	Same observation as above

2. RELIABILITY—Are data collection processes stable and consistent over time?

	Yes	No	Comments
➢ Is the same sampling method used from year to year, location to location, data source to data source?			Not applicable
10. Are similar currency conversions and inflation factors used as with other indicators or among the component parts of this indicator?		X	The data reported do not correct for inflation or, most significantly for exports, currency fluctuations (US dollar to the Euro) that significantly affected the reported exports.
Internal quality control			
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			No observed
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			No observed
➢ Do these procedures provide for periodic sampling and quality assessment of data?			No observed
➢ Other			
Transparency			
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?		X	
➢ Are data problems at each level reported to the next level?			Not applicable
➢ Are data quality problems clearly described in final reports?		X	In the recent Portfolio Review, for example, the indicators are reported in the wording of the indicator without distinction between the performance of the over sectors reported and the coverage of USAID assisted entities within each.

Findings:

- The data reported do not correct for inflation or, most significantly for exports, currency fluctuations (US dollar to the Euro) that significantly affected the reported exports.

Recommendations for improvement:

1. The SO team, together with the IP should figure a way to adjust export values reported in US dollars for currency fluctuations.

3. TIMELINESS—Are data collected frequently and are they current?

	Yes	No	Comments
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	X		
➢ Is a regularized schedule of data collection in place to meet program			Not observed

3. TIMELINESS—Are data collected frequently and are they current?

	Yes	No	Comments
management needs?			
➢ Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	X		
➢ Other			
Currency			
➢ Are the data reported in a given timeframe the most current practically available?	X		
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	X		
➢ Are the data reported as soon as possible after collection?			
➢ Is the date of collection clearly identified in the report?			

Findings:

Recommendations for improvement:

4. OBJECTIVITY—Can it be independently recognized, at "face value", as an acceptable measure of the desired result?

	Yes	No	Comments
➢ Is it unambiguous about what is being measured?	X		Except for the observations made above under "Validity"
➢ Is there general agreement over the interpretation of the results?	X		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	X		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	X		
➢ Other			

Recommendations for improvement:

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?

	Yes	No	Comments
➢ Are timely data available (i.e., is data current and available on regular basis)?	X		
➢ Can the data be collected frequently enough to inform management decisions?	X		
➢ Are the costs of data collection reasonable?	X		
➢ Are data gathered from secondary, e.g. government, sources?	X		
➢ If so, are data reliable, particularly in the way they are gathered and calculated?	X		
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?	X		
➢ Other			

Recommendations for improvement:

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?

	Yes	No	Comments
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?	X		It's an impact measure
➢ Does the indicator effectively tell the mission if it is "on track" regarding expected performance to-date in the implementation of the program plan?		X	This was discussed above under "validity"
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	X		
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		X	
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?		X	
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?	X		
➢ Impact of services	X		

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?

	Yes	No	Comments
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			

Findings:

- The intended result, i.e. movement in the exports of selected sectors through MCA cluster interventions, is a long term result. The indicator is similarly at too high a level to reflect short-term "on-trackedness".

Recommendations for improvement:

1. The SO teams should add indicators more at the level of process milestones and intermediate outputs.

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Enabling environment for investment improved

Performance indicator: Speed of business registration

Data source(s): Three Commercial Courts: Stip, Bitola and Skopje

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID officials participating or consulted: Jovan Madjovski and Zdravko Sami

Contractors: Checchi and Company Consulting Inc. – Harry Carr and Ilija Todorovski

1. VALIDITY—Do the data adequately represent performance?			
Face Validity	**Yes**	**No**	**Comments**
➢ Is there a solid, logical relation between the activity or program and what is being measured?	**X**		
➢ Does it closely measure the result it is intended to measure?	**X**		However, the indicator is based on the assumption that if the amount of time is reduced for registering a business then that should reflect the fact that administrative procedures have improved. However, elapsed time can greatly improve due to other causes, including "under-the-table" payment. The IP noted that changes in processing time due to administrative improvement will be small, requiring precise measurement.
➢ Can changes in the value of the indicator be reasonably attributed to the efforts of USAID and its partners?	**X**		Administrative changes can. However, as is described below under objectivity, elapsed time can also be reduced by "under-the-table" payoffs to court officials, particularly judges.
➢ Does the indicator effectively tell the mission if it is "on track" regarding expected performance to -date in the implementation of the program plan?	**X**		Measured at the current level of precision yes, it does this well. When aspects of the process are computerized it will tell "on-trackedness" even better.
Measurement Error			
Sampling Error (only applies when the data source is a survey)			Not applicable
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed up?			
Non Sampling Error			Not applicable
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
➢ What is the data transcription process? Is there potential for error?	**X**		Base data, necessary to the calculation of the indicator, are taken from hand written ledgers in three different courts. There is he potential for error in the original entry and there is potential for error in transcription, due to illegibility, incorrect reading, etc.
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)	**X**		The data gatherer is aware of the potential for error and is careful
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with missing data been correctly applied?			
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments

Findings:

- There is a slight potential for transcription error since the original data are recorded by hand in ledgers.
- The indicator is based on the assumption that if the amount of time is reduced for registering a business then that should reflect the fact that administrative procedures have improved. However, elapsed time can greatly improve due to other causes, including "under-the-table" payment. The IP noted that changes in processing time due to administrative improvement will be small, requiring precise measurement. Putting it another way, if there were not an opportunity, in business registration, for "rent seeking behavior" the time for processing registration would be reduced to a greater extent than improved administration.

Recommendations for improvement:

1. Although the potential for error is not particularly significant, the project CG & CL Project could explore the possibility of doing periodic checks from either the Commercial Registry of Businesses or the Public Revenues Office.
2. The IP carries the time reported to two decimal places in an attempt to capture changes with the necessary precision.

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?	**X**		
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?	**X**		The same method is used, although it is not written in a manual.
➢ Is the same sampling method used from year to year, location to location, data source to data source?			Not applicable
➢ Other			
Internal quality control			
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			Not observed See the recommendation under "validity" regarding reducing error.
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			Not observed
➢ Do these procedures provide for periodic sampling and quality assessment of data?			Not observed
➢ Other			
Transparency			
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?		**X**	They are not particularly onerous or complicated.
➢ Are data problems at each level reported to the next level?			Not applicable
➢ Are data quality problems clearly described in final reports?			Not observed
Recommendations for improvement:			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		
➢ Is a regularized schedule of data collection in place to meet program management needs?	**X**		
➢ Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	**X**		
➢ Other			
Currency			
➢ Are the data reported in a given timeframe the most current practically available?	**X**		
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	**X**		
➢ Are the data reported as soon as possible after collection?	**X**		
➢ Is the date of collection clearly identified in the report?			Not observed
Recommendations for improvement:			

4. OBJECTIVITY—Can it be independently recognized, at "face value", as an acceptable measure of the desired result?			
	Yes	**No**	**Comments**
➢ Is it unambiguous about what is being measured?	**X**		With the caution noted above that while it does measure administrative improvement. It also reflects favorable processing, "under-the-table" payments, etc.
➢ Is there general agreement over the interpretation of the results?	**X**		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	**X**		Except as noted above between administrative improvement and reduced "rent seekingness".
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	**X**		
➢ Other			
Recommendations for improvement:			

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	**No**	**Comments**
➢ Are timely data available (i.e., is data current and available on regular basis)?	**X**		
➢ Can the data be collected frequently enough to inform management decisions?	**X**		
➢ Are the costs of data collection reasonable?	**X**		
➢ Are data gathered from secondary, e.g. government, sources?	**X**		
➢ If so, are data reliable, particularly in the way they are gathered and calculated?	**X**		Except as noted above, regarding hand written entries in a ledger.
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?	**X**		
➢ Other			
Recommendations for improvement:			

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?	**X**		It is a process measure.
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	**X**		There are actually four indicators that measure the result.
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		**X**	
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?		**X**	
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?		**X**	
➢ Impact of services		**X**	
➢ Quality of services		**X**	

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
➢ Customer (end-user) satisfaction		**X**	
➢ Physical product		**X**	
➢ Major milestone		**X**	
➢ Other			
Recommendations for improvement:			

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Enabling environment for investment improved

Performance indicator: Shareholder awareness of their rights

Data source(s): Survey of shareholders

Partner or contractor who provided the data (if applicable): Commercial Law Project

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID:

Contractors: Checchi and Company Consulting Inc. – Harry Carr

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
Face Validity	**X**		
➢ Is there a solid, logical relation between the activity or program and what is being measured?			
➢ Other			
Measurement Error			
Sampling Error (only applies when the data source is a survey)			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
➢ Were samples representative?		**X**	There is some question of how much the baseline shareholder population is representative of the universe of shareholders.
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?	**X**		They were probably clear to shareholders but not clear to respondents who were not shareholders, but this should affect the validity of shareholder answers.
➢ If the instrument was self-reporting were adequate instructions provided?			Not applicable
➢ Were response rates sufficiently large?			Not applicable, however the issue may turn out to be whether there should have been over-sampling of the general population to ensure validity in the small number of respondents ending up making the shareholder baseline.
➢ Has non-response rate been followed up?			Not applicable
Non Sampling Error			
➢ Is the data collection instrument well designed?	**X**		
➢ Were there incentives for respondents to give incomplete or untruthful information?			Not observed
➢ Are definitions for data to be collected operationally precise?		**X**	The process of codifying open ended responses to a very general question into 8 specific categories of knowledge is a pretty sophisticated methodological thing to do and to train into enumerators. Recorded responses would very much vary between enumerators and, probably, for each enumerator from one respondent to another depending on the respondent's clarity of understanding and communication skills as well, on the enumerator side, by how well he/she was listening in each case or how tired they might have been.
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			Not observed
➢ Were there efforts to reduce the potential for personal bias by enumerators?			Not observed
➢ Other			
Transcription Error			
➢ What is the data transcription process? Is there potential for error?			Not observed
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			Not observed

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
➢ Have data errors been tracked to their original source and mistakes corrected?			Not observed
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?		**X**	The COP is unsure of the methods used. (See reliability)
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			Same as above
➢ Have procedures for dealing with missing data been correctly applied?			Same as above
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			It is impossible to tell as the method of calculation is unknown to the IP.
➢ Other			
Representativeness of Data			
➢ Is the sample from which the data are drawn representative of the population served by the activity?		**X**	See above
➢ Did all units of the population have an equal chance of being selected for the sample?			See above
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			See above
➢ Is the sample of adequate size?		**X**	See above. This probably is the main issue with how the baseline were gathered.
➢ Are the data complete? (i.e., have all data points been recorded?)			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**

Finding:

- The current baseline (2003) is from a random sample survey of the general population, taking from it those respondents who are shareholders. For the second year the IP plans to draw a sample of shareholders directly from a listing in the Commercial Register. There are several implications in this methodology change which could cast doubt on the comparability of future measurement of "shareholder awareness" *vis a vis* the baseline measure.
  - There is an assumption that the 139 shareholders drawn from the general population (baseline) survey are representative of the approximate 220,000 registered shareholders, specifically that they have a relatively low "awareness" level (29%, see reliability below).
  - Next year the real shareholder population will be asked the same question. What conclusions can reasonably be drawn from the number measuring "awareness" that is generated from that survey?
- Because of the method for distinguishing from verbal responses a very general open-ended question to any of eight specific categories of shareholder right (see above for more detail), whether a respondent ended up being categorized as knowledgeable about any of the 8 "rights" looked for would have a lot of "float and bounce" to it.

Recommendations for improvement:

1. The SO Team should be very sure that it understands all the implications of the change in survey methods from baseline to Year One iteration and is comfortable with the baseline drawn in 2003.
2. If there is uncertainty, the SO Team should decide to use the Year One measure, generated from the universe of real shareholders, as the new baseline.
3. In order to reduce the problem of the second finding above, the IP must:
   a. Train enumerators well
   b. Have as few enumerators as possible
   c. Be able to control circumstances of questioning to ensure careful attention of both the enumerator and the respondent.
4. A more ideal way to generate this information would be in writing, with respondents being in a controlled environment and given consistent instructions.

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?		**X**	The baseline data were collected from a random sample of the general population, as a sub-set of those respondents who were found to be shareholders. In Year Two the IP plans to draw its sample of shareholders from a listing now available at the Commercial Registry.
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?	**X**		That is the plan

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
➢ Is the same sampling method used from year to year, location to location, data source to data source?		**X**	The issue is described above.
11. Other			
Internal quality control			
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			Not observed
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			Not observed
➢ Do these procedures provide for periodic sampling and quality assessment of data?		**X**	It is a major survey effort.
➢ Other			
Transparency			
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?			Not observed The IP used two different local firms to do the data gathering. Presumably a manual was used but this was not verified.
➢ Are data problems at each level reported to the next level?			Not observed
➢ Are data quality problems clearly described in final reports?	**X**		A clear statement in at least one report says "These data must be qualified in that they represent the opinions of shareholders drawn from a larger survey of the general population, rather than from the census data of joint stock companies."

Findings:

➢ It is unclear how the baseline reported (29%) for 2003 was calculated. The project COP was unaware and has requested clarification from the HQ research specialist who made the calculation. Although the COP does intend to use the same question for the shareholders surveyed in 2004 it is absolutely necessary that he make the calculation in the same way that the baseline was made.

Recommendations for improvement:

1. The method for calculating "shareholder awareness" from the survey responses must be documented and put in the form of a manual in order to ensure consistent calculation.

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		
➢ Is a regularized schedule of data collection in place to meet program	**X**		The survey is done in September so as to accommodate the Mission's Annual Reporting

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
management needs?			requirement.
➢ Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	**X**		
➢ Other			
Currency			
➢ Are the data reported in a given timeframe the most current practically available?	**X**		
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	**X**		
➢ Are the data reported as soon as possible after collection?			Not observed
➢ Is the date of collection clearly identified in the report?	**X**		
Recommendations for improvement:			

4. OBJECTIVENESS—Can it be independently recognized, at "face value", as an acceptable measure of the desired result?			
	Yes	**No**	**Comments**
➢ Is it unambiguous about what is being measured?	**X**		
➢ Is there general agreement over the interpretation of the results?	**X**		
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?	**X**		
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?	**X**		
➢ Other			
Recommendations for improvement:			

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?			
	Yes	**No**	**Comments**
➢ Are timely data available (i.e., is data current and available on regular basis)?	**X**		
➢ Can the data be collected frequently		**X**	

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?

	Yes	No	Comments
enough to inform management decisions?			
➢ Are the costs of data collection reasonable?			The DQA team asked the IP for the costs of running the survey, both direct and from project overhead such as time of project staff, and he was not able to answer. There are certainly design saving, as he mentioned, that will be recognized in future years.
➢ Are data gathered from secondary, e.g. government, sources?			
➢ If so, are data reliable, particularly in the way they are gathered and calculated?			
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?			
➢ Other			

Findings:

➢ The DQA Team Leader asked the IP for the costs of running the survey, both direct and from project overhead such as time of project staff, and he was not able to answer. There are certainly design saving, as he mentioned, that will be recognized in future years. Moreover, the survey provides the IP with considerable more data than just this one indicator.

Recommendations for improvement:

1. The SO teams should ask the IP to provide an estimate of the costs so that at least it would have an idea of about the practicability of this indicator.

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?

	Yes	No	Comments
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?		X	
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	X		
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?		X	
➢ Is there an aspect of timing		X	

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
or timeliness, important to the desired result, that is missing?			
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?			
➢ Impact of services	**X**		
➢ Quality of services		**X**	
➢ Customer (end-user) satisfaction		**X**	
➢ Physical product		**X**	
➢ Major milestone			Not applicable
➢ Other			
Recommendations for improvement:			

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Enabling environment for investment improved

Performance indicator: Additional shareholders registered

Data source(s): Public Revenue Office

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID: Geoff Minott

Contractors: Checchi and Company Consulting Inc. – Harry Carr

1. VALIDITY—Do the data adequately represent performance?

Face Validity	Yes	No	Comments
➢ Is there a solid, logical relation between the activity or program and what is being measured?		X	It is a proxy and a good one. See below.
➢ Other			
➢ Is it a proxy?:	X		The indicator is measuring taxpaying entities (not individuals). As more of these are added to the government registry as taxpayers they represent a conscious choice to join the formal economic sector of Macedonia. They would do this, presumably but reasonably, only if the business environment were more attractive, conducive than the informal sector. Although this does not exactly reflect an investment decision, as the result seeks, it reflects probable anticipated investments.
➢ If so, is it as directly related to the relevant result as possible?	X		
➢ Does change in the proxy truly reflect change in the desired result?	X		It's pretty sure that experts would agree that this is a good proxy for the intended result.
➢ Is it tautological, i.e. is the "proxy" really the desired result itself in another form?		X	
Measurement Error			Not applicable
Sampling Error (only applies when the data source is a survey)			
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed up?			
Non Sampling Error			
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			Not applicable
➢ What is the data transcription			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments
process? Is there potential for error?			
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with missing data been correctly applied?			
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments

Findings:

- The indicator is a proxy. It is measuring taxpaying entities (not individuals). The logic is that as more of these are added to the government registry as taxpayers they represent a conscious choice to join the formal economic sector of Macedonia. They would do this, presumably but reasonably, only if the business environment were more attractive, conducive than the informal sector. Although this does not exactly reflect an investment decision, as the result seeks, it reflects probable anticipated investments.

Recommendations for improvement:

1. The SO team should be conscious that this is a proxy and watch it for any breakdown of its valid proxy properties. In its reporting the SO team should note it as a proxy.

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	No	Comments
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?	X		
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?	X		
➢ Is the same sampling method used from year to year, location to location, data source to data source?	X		
➢ Other			
Internal quality control			
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			Not observed
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			Not observed
➢ Do these procedures provide for periodic sampling and quality assessment of data?	X		
➢ Other			
Transparency			
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?			Not observed

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
➢ Are data problems at each level reported to the next level?			Not observed
➢ Are data quality problems clearly described in final reports?	**X**		
Recommendations for improvement:			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency			
➢ Are data available on a frequent enough basis to inform program management decisions?	**X**		
➢ Is a regularized schedule of data collection in place to meet program management needs?	**X**		
➢ Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest?	**X**		
➢ Other			
Currency			
➢ Are the data reported in a given timeframe the most current practically available?	**X**		
➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?)	**X**		
➢ Are the data reported as soon as possible after collection?	**X**		
➢ Is the date of collection clearly identified in the report?			Not observed
Recommendations for improvement:			

4. OBJECTIVITY—Can it be independently recognized, at "face value", as an acceptable measure of the desired result?			
	Yes	**No**	**Comments**
➢ Is it unambiguous about what is being measured?		**X**	First, the wording implies that individual taxpayers are being counted.
➢ Is there general agreement over the			

4. OBJECTIVITY—Can it be independently recognized, at "face value", as an acceptable measure of the desired result?

	Yes	No	Comments
interpretation of the results?			
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?			
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?		X	
➢ Other			

Findings:

- The wording of the indicator is too vague to show what it really is measuring, i.e. taxpaying entities, not individuals.
- The variety of "taxpaying entities" is staggering and ranged from government agencies to casinos and includes such things associations, political parties and municipalities. At the same time it does not include sole proprietors or partnerships. There doesn't seem to be a significant problem with the breadth of the indicator as it is operationally defined. But the range should be explicit to the SO team and in the records.

Recommendations for improvement:

1. Change the wording so that it is clear that this is measuring new tax paying entities, most of which are businesses and all of which pay some form of business related tax, i.e. employee withholding, VAT, excise, profit.
2. The SO team should consider all the things that are captured (and not captured) in the indicator and decide if this is the operational definition desired.

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?

	Yes	No	Comments
➢ Are timely data available (i.e., is data current and available on regular basis)?	X		
➢ Can the data be collected frequently enough to inform management decisions?	X		
➢ Are the costs of data collection reasonable?	X		
➢ Are data gathered from secondary, e.g. government, sources?	X		
➢ If so, are data reliable, particularly in the way they are gathered and calculated?	X		
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?	X		
➢ Other			

Recommendations for improvement:

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?	**X**		
➢ Taken as a group, are the indicator and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?	**X**		This one of four indicators measuring the result in question.
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?			Not applicable, as this is a proxy.
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?			
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?			Not applicable, as this is a proxy.
➢ Impact of services			
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			
Recommendations for improvement:			

DATA QUALITY ASSESSMENT CHECKLIST

Strategic Objective: Accelerated Development and Growth of Private Sector

Intermediate Result: Enabling environment for investment improved

Performance indicator: Compliance to WTO requirements

Data source(s): WTO Compliance Project

Partner or contractor who provided the data (if applicable):

Is this indicator reported in the Annual Report?

Date(s) of assessment: June, 2004

Location(s) of assessment: Skopje, Macedonia

Assessment team members:

USAID: Zdrakvo Sami

Contractors: Checchi and Company Consulting Inc. – Harry Carr

1. VALIDITY—Do the data adequately represent performance?			
Face Validity	**Yes**	**No**	**Comments**
➢ Is there a solid, logical relation between the activity or program and what is being measured?	**X**		On the surface there is face validity between this indicator and the result it is intended to measure. However, it is not an index but rather four separate indicators that cannot be aggregated to yield one measure. Four things are tracked: • WTO laws enacted • Notifications submitted • Governmental coordinating mechanism meeting held and • Topic analysis reports published as a measure of institutional capacity for trade analysis
Directness			
➢ Does it closely measure the result it is intended to measure?	**X**		Conceptually a unified measure or index would be a very direct measure of the intended result – "Enabling environment for investment improved"
➢ Is it grounded in theory and practice?	**X**		The SO Team should find out whether here exists such a measure. WTO has been such an important process for so many years and the subject of work by USAID missions that there must be some generally accepted measure.
➢ Does it represent an acceptable measure to both proponents and skeptics?			
➢ Other aspects of directness			
Measurement Error			Not applicable
Sampling Error (only applies when the data source is a survey)			
➢ Were samples representative?			
➢ Were the questions in the survey/questionnaire clear, direct, easy to understand?			
➢ If the instrument was self-reporting were adequate instructions provided?			
➢ Were response rates sufficiently large?			
➢ Has non-response rate been followed up?			
Non Sampling Error			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	**No**	**Comments**
➢ Is the data collection instrument well designed?			
➢ Were there incentives for respondents to give incomplete or untruthful information?			
➢ Are definitions for data to be collected operationally precise?			
➢ Are enumerators well trained? How were they trained? Were they insiders or outsiders? Was there any quality control in the selection process?			
➢ Were there efforts to reduce the potential for personal bias by enumerators?			
➢ Other			
Transcription Error			Not applicable
➢ What is the data transcription process? Is there potential for error?			
➢ Are steps being taken to limit transcription error? (e.g., double keying of data for large surveys, electronic edit checking program to clean data, random checks of partner data entered by supervisors)			
➢ Have data errors been tracked to their original source and mistakes corrected?			
➢ If raw data need to be manipulated to produce the data required for the indicator:			
➢ Are the correct formulae being applied?			
➢ Are the same formulae applied consistently from year to year, site to site, data source to data source (if data from multiple sources need to be aggregated)?			
➢ Have procedures for dealing with missing data been correctly applied?			
➢ Are final numbers reported accurate? (E.g., does a number reported as a "total" actually add up?)			
➢ Other			
Representativeness of Data			Not applicable
➢ Is the sample from which the data are drawn representative of the population served by the activity?			
➢ Did all units of the population have an equal chance of being selected for the sample?			
➢ Is the sampling frame (i.e., the list of units in the target population) up to date? Comprehensive? Mutually			

1. VALIDITY—Do the data adequately represent performance?			
	Yes	No	Comments
exclusive (for geographic frames)			
➢ Is the sample of adequate size?			
➢ Are the data complete? (i.e., have all data points been recorded?)			

Findings:

- The indicator is not an index, as shown in the PMP, but rather four separate indicators that cannot be aggregated to yield one measure. Four things are tracked:
  - WTO laws enacted
  - Notifications submitted
  - Governmental coordinating mechanism meeting held and
  - Topic analysis reports published as a measure of institutional capacity for trade analysis

Recommendations for improvement:

1. The SO Team, together with the IP, should develop or find an exiting milestone measure of compliance. Possibly looking at the experience of other countries or in WTO methods for performance measurement some form of standardized or generally accepted milestone measure could be found.

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	No	Comments
Consistency			
➢ Is a consistent data collection process used from year to year, location to location, data source to data source (if data come from different sources)?			
➢ Is the same instrument used to collect data from year to year, location to location? If data come from different sources are the instruments similar enough that the reliability of the data are not compromised?			
➢ Is the same sampling method used from year to year, location to location, data source to data source?			
➢ Other			
Internal quality control			
➢ Are there procedures to ensure that data are free of significant error and that bias is not introduced?			
➢ Are there procedures in place for periodic review of data collection, maintenance, and processing?			
➢ Do these procedures provide for periodic sampling and quality assessment of data?			
➢ Other			
Transparency			

2. RELIABILITY—Are data collection processes stable and consistent over time?			
	Yes	**No**	**Comments**
➢ Are data collection, cleaning, analysis, reporting, and quality assessment procedures documented in writing?			
➢ Are data problems at each level reported to the next level?			
➢ Are data quality problems clearly described in final reports?			
Recommendations for improvement:			

3. TIMELINESS—Are data collected frequently and are they current?			
	Yes	**No**	**Comments**
Frequency ➢ Are data available on a frequent enough basis to inform program management decisions? ➢ Is a regularized schedule of data collection in place to meet program management needs? ➢ Are data collected at a time in the year appropriate to the desired meaning of the indicator, e.g. agriculture produce pricing *vis a vis* harvest? 12. Other **Currency** ➢ Are the data reported in a given timeframe the most current practically available? ➢ Are data from within the policy period of interest? (i.e., are data from a point in time after intervention has begun?) ➢ Are the data reported as soon as possible after collection? ➢ Is the date of collection clearly identified in the report?			
Recommendations for improvement:			

4. OBJECTIVENESS—Can it be independently recognized, at "face value", as an acceptable measure of the desired result?			
	Yes	**No**	**Comments**
➢ Is it unambiguous about what is being measured?			

4. OBJECTIVENESS—Can it be independently recognized, at "face value", as an acceptable measure of the desired result?

	Yes	No	Comments
➢ Is there general agreement over the interpretation of the results?			
➢ Is it unidimensional (i.e., does it measure only one phenomenon at a time)?			
➢ Is it operationally precise (i.e., is there no ambiguity over what kind of data should be collected)?			
➢ Other			

Recommendations for improvement:

5. PRACTICALITY—Can data be obtained at a reasonable cost and timely fashion?

	Yes	No	Comments
➢ Are timely data available (i.e., is data current and available on regular basis)?			
➢ Can the data be collected frequently enough to inform management decisions?			
➢ Are the costs of data collection reasonable?			
➢ Are data gathered from secondary, e.g. government, sources?			
➢ If so, are data reliable, particularly in the way they are gathered and calculated?			
➢ If so, are data valid, e.g. are they really measuring what they purport to measure?			
➢ Other			

Recommendations for improvement:

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?

	Yes	No	Comments
➢ Does it merely indicate progress rather than attempt to fully describe everything an activity accomplishes?			
➢ Taken as a group, are the indicator			

6. ADEQUACY—How completely does the indicator, together with its companion indicators if more than one, measure the desired result?			
	Yes	**No**	**Comments**
and its companion indicators the minimum necessary to ensure that progress toward the intended result is sufficiently captured?			
➢ Is there a particularly important dimension to the result that is missed in the indicator or group of indicators?			
➢ Is there an aspect of timing or timeliness, important to the desired result, that is missing?			
➢ Other aspect of "sufficiency?			
➢ Does the indicator reflect an outcome of the program, as opposed to the completion of an activity or process?			
➢ Impact of services			
➢ Quality of services			
➢ Customer (end-user) satisfaction			
➢ Physical product			
➢ Major milestone			
➢ Other			
Recommendations for improvement:			

ATTACHMENT D

Resumes of the Macedonian Research Specialists

CURRICULUM VITAE

NIKOLINA KENIG Born: December 19, 1968

Education:

University of *Sts. Cyril and Methodius*, Republic of Macedonia	**(1992) B.A. Psychology (GPA 4.00)**
University of *Notre Dame*, Indiana	**(1994) M.A. International Peace Studies (GPA 3,95)**
University of *Sts. Cyril and Methodius*, Republic of Macedonia	**(2000) M.A. Social Psychology (GPA 4.00)**

Employment:

1995 ---	University Assistant of **Psychometrics, Methodology of Psychological Research** and Psychology of Conflict Resolution at Faculty of Philosophy (University of *Sts. Cyril and Methodius* - Skopje)
1996 ---	**Ethnic Conflict Resolution Project** (senior assistant, since 2000 - Acting Director) - Center for Human Rights and Conflict Resolution, Institute for Sociological, Political and Juridical Research

Study Grants:

1992 -1993	Ministry of Science of the Republic of Macedonia (M.A. studies in Social Psychology, University of Sts. Cyril and Methodius – Skopje, R. of Macedonia)
1993-1994	*NAFSA* (M.A. studies in International Peace Studies, University of Notre Dame, Indiana, USA)
1999	*Tempus PHARE* (Education in Advanced Statistical Methods of Data Analysis, ESADE, University of Ramon Llull – Barcelona, Spain)

Awards:

1992	Manning Award
1996	University award for development and promotion of the Faculty

Key qualifications:

- measurement, evaluation, and methodology of research in social and behavioural sciences

- psychological aspects of inter-ethnic relations; conflict resolution; nationalism and gender

Research Projects:

1997 --	**Experimental Teaching Project** (project consultant). Ministry of Education and Physical Culture of the Republic of Macedonia
1997--	**European Network of Health Promoting Schools** (member of the expert team). Pedagogical Institute and the Center for Psychosocial Crisis Intervention
1997--	**Ethnic Stereotypes among Future Pre-school and Primary-school Teachers** (researcher), Center for Human Rights and Conflict Resolution
1998-00	**Ethnic Identities of Macedonians and Albanians in the Republic of Macedonia and the Possibility to Develop Civic-oriented Identity** (researcher), Center for Human Rights and Conflict Resolution
2000-01	**Gender Identity of Macedonian Women in different Cultural Contexts - Parents and their Daughters,** (director), Eurobalkan-International Center for European Culture
2001--	**Our Neighborhood – Educational TV Show** (assistant, senior researcher since 2004). Ethnic Conflict Resolution Project and Center for Human Rights and Conflict Resolution
2004--	**Cultural Dimensions of the ethnic Macedonian and ethnic Albanian communities in the Republic of Macedonia** – doctoral research

Additional education

1992-2001	Degree in Applied Gestalt Therapy - Gestalt Center - Agency for Psychological and Psychotherapeuthical Education - Belgrade
1999, July	*The Image of The Other*, Open Society Institute, Skopje , International Summer School, (participant)
2000, May-June	Supervised education in Advanced Statistical Methods of Data Analysis, ESADE, University of Ramon Llull, Barcelona (researcher)

2001, 18th -29th June	Teaching the Process of Inquiry in Psychological Science, summer school, IUC Dubrovnik, (participant)
2002, 24th June-12th July	Methodology of Cross-Cultural Research, Psychological Science, summer school, IUC Dubrovnik, (participant)
2003, 10th January -6th February	Guest leturer (Women on the Balkans) and individual research in nationalism and gender issues, Arizona State University, Tempe, USA

Relevant activities

1998, 1999, January	Peace Theories, Conflict and Conflict Resolution, *series of courses for graduate students* at the Balkans Peace Studies, Macedonia, (lecturer)
1999 - 2000	Mediation at Schools, series of workshops for training mediators, ECRP (supervisor)
2002, June-August	External evaluation of *Post Conflict Confidence Building Programme in the Republic of Macedonia, International Alliances - Holland* (evaluator)
2003, June 13th -16th and *November 5th - 7th*	Methodology of scientific research, *summer school for students* sponsored by Romaversitas, Macedonia (lecturer)

Anica Dragovik, M.A

Institute for Sociological, Political and Juridical Research,

Bulevar "Partizancki odredi" No.89, Skopje, Republic of Macedonia,

Phone (389 2) 3061 119, Fax (389 2) 3061 282, mobile: 070 305 174

E-mail: anica@isppi.ukim.edu.mk ; a_dragovic@yahoo.com

CURRICULUM VITAE

Date of Birth: 07.08. (August), 1966, Tetovo, Republic of Macedonia

Citizenship: Republic of Macedonia

Working Experience: 10 years

Education:

- Enrolled in the PH.D studies, Faculty of Philosophy Department of Sociology at the University "Sts. Ciril and Methodius", Skopje.
- Master of Philosophy in Demography, (2003), title: "Recent Levels, Differentials and Fertility in the Republic of Macedonia, Cairo Demographic Center, Cairo, Egypt.
- Special Diploma in Population and Development, (2001), title: The Impact of Women's Status on Fertility in the Republic of Macedonia, Cairo Demographic Center, Cairo, Egypt.
- General Diploma in Demography, (2000), title: "Level, Trend and Differential Fertility in the Republic of Macedonia", Cairo Demographic Center, Cairo, Egypt, 2000.
- Master of Art (1997), title: "Macro-sociological Analysis of the Parliamentary Elections in the Central and Eastern Europe: 1989-1994", Faculty of Philosophy Department of Sociology at the University "Sts. Ciril and Methodius", Skopje, 1997.
- Faculty of Philosophy, Department of Sociology at the University "St. Cyril and Methodius", (1992) Skopje.

Training and Professional Courses:

Attended successfully during the Academic Years 2000-2003 year at the Cairo

Demographic Center, Cairo, Egypt. Beside the main topics related to the work in the fields of demography (population), provided with educational opportunities to study supporting course in the following topics: Statistics 75 hours, Social Research Methods, Advanced Research Techniques and Methodologies and Statistical Methods, and Computer Programs, basically using software packages for statistical analysis in social data.

Language Skills:

Read, Speak and write: Macedonian, Serbian and English language.

Other Skills:

Know to use and work in Word, Excel, PDP, SPSS, and AMOS 4 programs.

Professional Interest:

Sociology, Demography, Social Research Method and Statistics for Social Science, and Soicolgy of Politic

Professional/Occupational Experience:

At the Institute for Sociological, Political and Juridical Research at the Sts. Cyril and Methodus University – Skopje, Republic of Macedonia, since 1993

- 1998 Research Assistant, Institute for Sociological, Political and Juridical Research
- 1993-1998, Junior Research Assistant Institute for Sociological, Political and Juridical Research

Research Experience:

Projects:

- "Project for the Analysis of Land Tenure and Agricultural Productivity in the Republic of Macedonia", conduct by the University of Wisconsin, USA, Land Tenure Center in Skopje and the Government of the Republic of Macedonia, February - May 1996; *Short time activity, as a interwier in a case study and a writer*

- "Public Opinion on Local Elections in Macedonia '96", Conduct by Institute for Sociological, Political and Juridical Research, Skopje, September 1996. *Statistical analysis*

- "Public Opinion on the Social an Economic Reforms", conducted by the Institute of Sociological, Political and Juridical research, Skopje, January-February,1997. *Preparing questions for conducting a poll, statistical analysis and report*

- "Local Elections in Macedonia '96", conducted by the Institute for Sociological, Political and Juridical Research, Skopje, January - April, 1997. *Preparing questions for conducting a poll, statistical analysis and report*

- "Function of the Local Government in Republic of Macedonia", conducted by the Institute for Sociological, Political and Juridical Research, Skopje, July 1997 - Jun 2000. *As a secretary of the Project*

"Public Opinion on Social an Economic Reforms", conducted by the Institute for Sociological, Political and Juridical Research, Skopje, October-Decembar,1997. *Preparing questions for conducting a poll, statistical analysis and report*

" Qualitative Study of the Living Standards of the Population epublic of Macedonia", conducted by World Bank, December 1997-Mart 1998. *As a interwer and writer of regional report*

"Sociological Aspects of the City Skopje"**,** conducted by Institute for Sociological, Political and Juridical Research, Skopje, February - Juny,1998, *As a writer*

"Public Opinion on Social an Economic Reforms", conducted by the Institute for Sociological, Political and Juridical research, Skopje, May-Juny,1998. *Preparing questions for conducting a poll, statistical analysis and report*

"Public Opinion on Social an Economic Reforms", conducted by the Institute of Sociological, Political and Juridical research, Skopje, May-Juny,1998. *Preparing questions for conducting a pol, statistical analysis and report*

"Public Opinion on Social an Economic Reforms", conducted by the Institute of Sociological, Political and Juridical Research, Skopje, May-Juny,1999. *Preparing questions for conducting a poll, statistical analysis and report*

"Local Government in the Republic of Macedonia", *As a member of the team and secretary of the Project (1999-2002)*

"*Sociological Aspects of the City Skopje*" survey conducted by Institute of sociological, political and juridical research, Skopje, February - Juny,1998, As a writer

"Rapid Beneficiary Assessment for the Basic and Secondary Education and Teacher Training",(2003) conducted by Ministry of Education and Science of Republic of Macedonia. As a Consultant, and data processing and statistical analysis

"Evaluation of Confidence Building Projects", (2003) conducted by USAID, office in Skopje. As a Consultant, and data processing and statistical analysis

"*Mapping the Socio-economic Disparities among Macedonian Municipalities*" (2003-2004), as an National consultant for the preparation of the Assessment Report – Publication conducted by United Nations Development Programme (UNDP),
and Government of the Republic of Macedonia

"Development of Municipalities and Culture - Social evaluation and evaluation of needs, research of local heritage: debar, Centar Zupa and Rostuse" (2004) conducted by the Ministry of Culture and world Bank

***"** Interrelation between Gender, Ethnicity and Fertility: The Significance of the Education and Employment on Fertility among different ethnic groups in the Republic of Macedonia" 2004, Cooperative Research Project, Faculty of Phosophy - Institute for Sociology and Foundation Open Society InstituteMacedonia, As a member of a team.*

"*Education Modernization Projec*t" (2004), conudcted by the Ministry of Education and Science in the Republic of Macedonia and the World Bank, as an consultant on the baseline data on access, quality and effectiveness for Improving Education Quality and Participation

Lecturing/Teaching experience:

University Courses/and level

- Methodology of Social Science - Quantitative Analysis", Postgraduate Course at the Institute for Sociological, Political and Juridical Research, Skopje
- Social Analysis using Computer Software, Postgraduate Course at the Institute for Sociological, Political and Juridical Research, Skopje

Publications:

- "Public Opinion on Social and Economic Reforms" 1997, 1998, 1999. Institute for Sociological, Political an Juridical Research, Skopje
- "Sociological Aspects of Skopje".1998 Institute for Sociological, Political an Juridical Research, Skopje
- "Local elections 1996", (1998) Institute for Sociological, Political an Juridical Research, Skopje
- "Political Culture of the Citizens and their Influence on the Local Government", (2002) Institute for Sociological, Political an Juridical Research, Skopje
- "Level, Pattern and Differentials of Fertility in the Republic of Macedonia, 2000, Population and Sustainable Development Research Monograph, Cairo Demographic Center, 2000, Cairo, Egypt.
- "Differentials of Fertility in the Republic of Macedonia". 2002. in New Balkans Politics, vol.2-3, Peace and Democracy Center, Skopje
- **"**Women's Status and Fertility in the Republic of Macedonia", Monograph, in Population and Sustainable Development Research Monograph, Cairo Demographic Center, 2001, Cairo, Egypt.

CURRICULUM VITAE

1. Surname : Todorovski
2. Name : Ilija
3. Date of birth : August 31, 1952
4. Nationality : Macedonian
5. Civil status : Married
6. Education :

Institution [Date from - Date to]	Degree(s) or Diploma(s) obtained:
Law Faculty – Skopje, R.Macedonia 1971 Oct – 1978 Nov	B.A. in law
Philological Faculty/Major: English Language and Literature, Skopje, R.Macedonia 1971Oct – 1992 June	Professor in English Language and Literature /B.A./
Institute for Sociological, Political and Juridical Research – Skopje, R.Macedonia	M.A. in Political sciences
Law Faculty – Skopje, R. Macedonia	Ph.D. in Political sciences

1. **Language skills:** Indicate competence on a scale of 1 to 5 (1 - excellent; 5 - basic)

Language	Reading	Speaking	Writing
English	1	1	1
Macedonian (native)	1	1	1
Serbian	1	1	1
Croatian	1	1	1
Bulgarian	1	2	2

2. **Membership of professional bodies:**

1. Group of Independent Experts on the European Charter of Local Self-Government at the Council of Europe, Strasbourg
2. Team of Local Self-Government Experts of the Regional German Foundation "Friedrich Ebert", Zagreb
3. **Team of the TEMPUS Programme of the European Union for Establishment of European Studies Faculty in Skopje**

3. **Other skills:** (e.g. Computer literacy, etc.)

Microsoft Word, Excel

4. **Present position:**

Institute for Sociological, Political and Juridical Research

1. Political scientist – Senior research fellow
2. Professor at Local Government at the Graduate Studies of the Institute

5. **Years within the firm:**

1979 – 2004

6. **Key qualifications:**(Relevant to the project)

Empirical research in all stages (executor, project team member, team leader) dealing with various aspects and problems of local government functioning:

- making interviews with ministers, mayors, councillors, citizens, etc
- participation in seminars as lecturer, seminar organizer, etc.
- working on local government legislation, etc.

v

7. **Specific experience in the region:**

Country	Date from - Date to

8. **Professional experience**

Date	Location	Company	Position	Description
1979-March - 2004 April	Skopje, R. Macedonia	Institutre for Sociological, Political and Juridical Research	Political scientist – researcher	Empirical research in the field of political sciences, functioning of the political system at central level, local level, conflicts, etc.
1995 Oct – 1997 March	Skopje, R. Macedonia	UNPREDEP (UN Preventive Deployment)	Media analyst	Making analyses in the field of political sciences

9. **Other relevant information:** (e.g., Publications)

SCIENTIFIC PAPERS /selected/:

- *Public Services at the Local Level in the Republic of Macedonia,* pp. 81-87, as a part of the anthology "Reforms of Public Services", Friedrich Ebert Stiftung, Zagreb, 2003, 100 pp.

- *Research Concept Supportive to the Creation and Implementation of the NATIONAL STRATEGY FOR SUSTAINABLE DEVELOPMENT OF THE REPUBLIC OF MACEDONIA,* Ministry of Environment

and Physical Planning and Institute for Sociological, Political and Juridical Research, Skopje, 2003, 209 pp. (as a co-author)

- *Structure and Relations between Executive and Legislative Bodies at Local Level in Macedonia,* pp. 7-21, as a part of the anthology "Executive and Legislature at Local Level", Friedrich Ebert Stiftung, Zagreb, 2002, 134 pp.

- *Legal and Social Status of the Minorities at Local level in the Republic of Macedonia,* pp. 32-46, as a part of the anthology "National Minorities in South-East Euirope", Friedrich Ebert Stiftung, Zagreb, 2002, 130 pp.

- *National Human Development Report 2001 /Social Exclusion and Human Insecurity in FYR Macedonia/,* UNDP, Skopje, 2002 (as a co-author)

- *Local Self-Government and Decentralization in the Republic of Macedonia,* pp. 172-193, as a part of the anthology "Local Self-Government and Decentralization in South-East Europe", Friedrich Ebert Stiftung, Zagreb, 2001, 272 pp.

- *Causes and the Social Context of the Insecurity of the Citizens in the Republic of Macedonia*, pp. 329-339, as a part of the anthology "Macedonia 1989-1999", Open Society Institute Macedonia, Skopje, 2001, 419 pp.

- *Local Government Reform Project - interpretation of the survey data*, Institute for Sociological, Political and Juridical Research, Skopje, 2001, 36 pp.

- *Local Government in Macedonia,* pp. 241-288, as a part of the anthology "Stabilization of Local Governments", Open Society Institute, Budapest, 2001, 472 pp.

- *Institutional, Organizational and Motivation Components of the Social Security System and Factors Influencing It*, pp. 167-222 (with a co-author), as a part of the study "Socio-economic Structure and Problems of the Population in the Republic of Macedonia", Institute for Sociological, Political and Juridical Research, Skopje ISPPI, Skopje, 2001, 250 pp.

- *Socio-economic Report on HEP "Chebren",* Ministry of Construction and Urban Planning, Skopje, 2000. 35 pp. (as a co-author)

- *Existing Financing of Local Government - Obstacle towards Its Development*, 123 (journal), Skopje, No. 5-6, January/February, 1999, str. 8-11

- *The Impact of the State upon the Economic Efficiency of Enterprises,* NIP Globus, Skopje, 1997, 252 pp.

- *Local Government in England, USA and Yugoslavia*, Studentski zbor, Skopje, 1991, 165 pp.

- *Local Government Problems in Macedonia's Newly-Established Municipalities*, Institute for Sociological, Political and Juridical Research, Skopje, 1999, 86 pp. (as a co-author

- *Institutional, Organizational and Motivational Components of the Social Security System and Factors Influencing It,* p. 181 - 235 *(*with a co-author); as a part of the study "*Socioeconomic Structure and Problems of the Population of the Republic of Macedonia",* Institute for Sociological, Political and Juridical Research, Skopje, 1999, 258 pp.

- *Political, Institutional and Organizational Components of Rural Revitalization in the Republic of Macedonia,* p. 81-98, as a part of the study "*Rural Revitalization in Macedonia",* Economics Institute & Institute for Sociological, Political and Juridical Research, Skopje, 1997, 199 pp.

- *Structure and Operation of the Macedonian Local Government,* Skopje, 1997, 20 pp. National Report to the EU Local Government Commission

- *Local Government in Macedonia***,** p. 129-143 as a part of the study *"Local Governments in Central and East Europe and the Community of the Independent States",* Institute for Local Government and Public Service, Budapest, Hungary, 1994, 227 pp.

- *Comparative Aspects of the Macedonian, Western and East-European Local Governments",* Pogledi, Skopje, 1995, p. 55-73

-*Efficiency of the Macedonian Political System*, p. 122-138, as a part of the study *"Efficiency of the Parliamentarian Democracy in Macedonia",* Institute for Sociological, Political and Juridical Research, Skopje, 1995, 215 pp.

-*Local Government in England, USA and Yugoslavia,* Studentski zbor, Skopje, 1991, 165 pp.

- *Social System and Strikes* (with co-authors), Institute for Sociological, Political and Juridical Research, Skopje, 1991, 293 pp.

RESEARCH EXPERIENCE (Projects - selected)

- *STRATEGY OF SUSTAINABLE DEVELOPMENT OF THE REPUBLIC OF MACEDONIA*
Research activities:
- Project co-ordination
-Participating in designing the general project idea;
-Writing a paper

- *NATIONAL HUMAN DEVELOPMENT REPORT 2001 /*Topic: *Social Exclusion and Human Insecurity in FYR Macedonia/*

Research activities:
- Project co-ordination
- Participating in designing the general project idea;
- Participating in making field research instruments;
- Collecting and analysis of documents and legal regulation dealing with the above topic;
- Identification the reasons for the negative developing tendencies in the field of political system and labour relations
- Proposing solutions and measures to improve the situation in the fields

- *FUNCTIONING OF THE LOCAL GOVERNMENT SYSTEM*

Research activities:

- Project co-ordination
- Designing the general project idea;
- Making field research instruments;
- Interviewing officials in the Macedonian Government, municipalities, local government associations and agencies, etc.
- Collecting and analysis of documents and legal regulation dealing with the local government system in Macedonia;
- Evaluation of the efficacy and efficiency of the local government institutions.

- *THE SOCIO-ECONOMIC CHARACTERISTICS AND PROBLEMS OF THE POPULATION IN THE REPUBLIC OF MACEDONIA*

Research activities:

- Designing a specific project idea;
- Participation in creation of the field research instruments;
- Interviewing officials in the Macedonian Parliament, Ministry of Labour and Social Policy, Ministry of Education, leaders of the major political parties, non-governmental organizations, social security institutions and agencies, etc.
- Collecting and analysis of documents and legal regulation dealing with the network of welfare institutions and population policy in Macedonia;
- Evaluation of the range and efficiency of the welfare (social security) system and population policy in Macedonia

- *THE SCIENTIFIC DEVELOPMENT IN THE REPUBLIC OF MACEDONIA*

Research activities

- Designing a specific project idea;
- Participation in creation of the field research instruments;
- Analyzing documents and statistical data on the situation in the natural sciences
- Evaluation of the developing tendencies in the field of natural sciences
- Proposing solutions and measures to improve the situation in the field

- *THE STRATEGY AND POLICY OF RURAL DEVELOPMENT IN MACEDONIA*

Research activities

- Designing a specific project idea;
- Participation in creation of the field research instruments;
- Interviewing officials in the Ministry of Agriculture and Ministry of Development, experts in agricultural and veterinarian institutes, municipal agricultural centers, etc
- Analyzing laws and other documents dealing with the status of agriculture, setting, organization, financing and work motivation of the agricultural institutes and centres, their infrastructure etc
- Evaluation of the rural development
- Proposing solutions and measures to improve the situation in the respective area

- *THE IMPACT OF EDUCATION UPON INDUSTRIALIZATION IN THE REPUBLIC OF MACEDONIA*

Research activities

- Designing a specific project idea;
- Participation in creation of the field research instruments;
- Interviewing officials in the Ministry of Education, experts in the Pedagogical Institute and secondary and tertiary (University) professors

- Analyzing laws and other documents dealing with the status of professional education in Macedonia, methods of work, organization, financing, staff motivation etc. at the secondary and tertiary schools and institutions dealing with secondary school curricula
- Evaluation of the professional education system efficiency
- Proposing solutions and measures to improve the situation in the respective area

- THE EFFICIENCY OF THE PARLIAMENTARIAN DEMOCRACY IN MACEDONIA

Research activities
- Designing a specific project idea;
- Participation in creation of the field research instruments;
- Interviewing members of Parliament and some Ministers
- Analyzing the procedure and participation of the Assembly and Government in drafting, passing and implementing laws
- Evaluation of the Macedonian political system efficiency
- Giving opinion to raise the political system efficiency

- LABOUR STRIKES IN MACEDONIA

Research activities
- Designing a specific project idea;
- Participation in creation of the field research instruments;
- Interviewing political and party officials, managers and experts in the enterprises and strikers
- Analyzing economic and labour laws regulating the status of the enterprises
- Analyzing the regulation of enterprises, their production indicators, work organization etc
- Revealing the profound causes of strikes
- Proposing solutions towards elimination of many governmental restrictions to enterprises

PARTICIPATION AT SCIENTIFIC GATHERINGS

(since 1999 in reverse order)

-Local Government Reforms in the Republic of Macedonia, Ohrid, 2004

-Reforms of Public Services, Budapest, 2003

-National Minorities in South-East Europe, Sarajevo, 2002

- Preliminary draft agenda for the second meeting of the Group of Independent Experts on the European Charter of Local Self-Government, Strasbourg, 2002

- Cross Border Cooperation in the Triangle Nish-Skopje - Sofija, Sofija 2002

- Decentralizing Government, Ljubljana 2002

- Local Self-government and Decentralization, Strasbourg, 2001

- Local Self-government and Decentralization, Zagreb, 2001

- Man and Environment - conditions for stable society, Skopje, 2001

- Local Government Reform, Skopje, 2000

- Crossborder Co-operation Sandanski - Strumica, Sandanski (Bulgaria) - Strumica (Macedonia), 2000

- Neighbourhood Services in Disadvantaged Urban Areas and in Areas of Low Population, Strasbourg , France, 2000

- Aspects of Local Government Decentralization, Debar (Macedonia), 2000

- Democratic Development in Denmark and the Republic of Macedonia, Denmark, 1999

- Decentralization and Government, Poland, 1999

CONSULTANCY

TRAINING AND CAPACITY BUILDING OF LOCAL SELF-GOVERNMENT INSTITUTIONS, European Agency for Reconstruction, 2003

USAID LOCAL GOVERNMENT PROJECT ASSESSMENT, USAID, 2003