Challenges in Improving Information Quality

NISS Data Quality Conference
November 30 – December 1

Ann Thornton

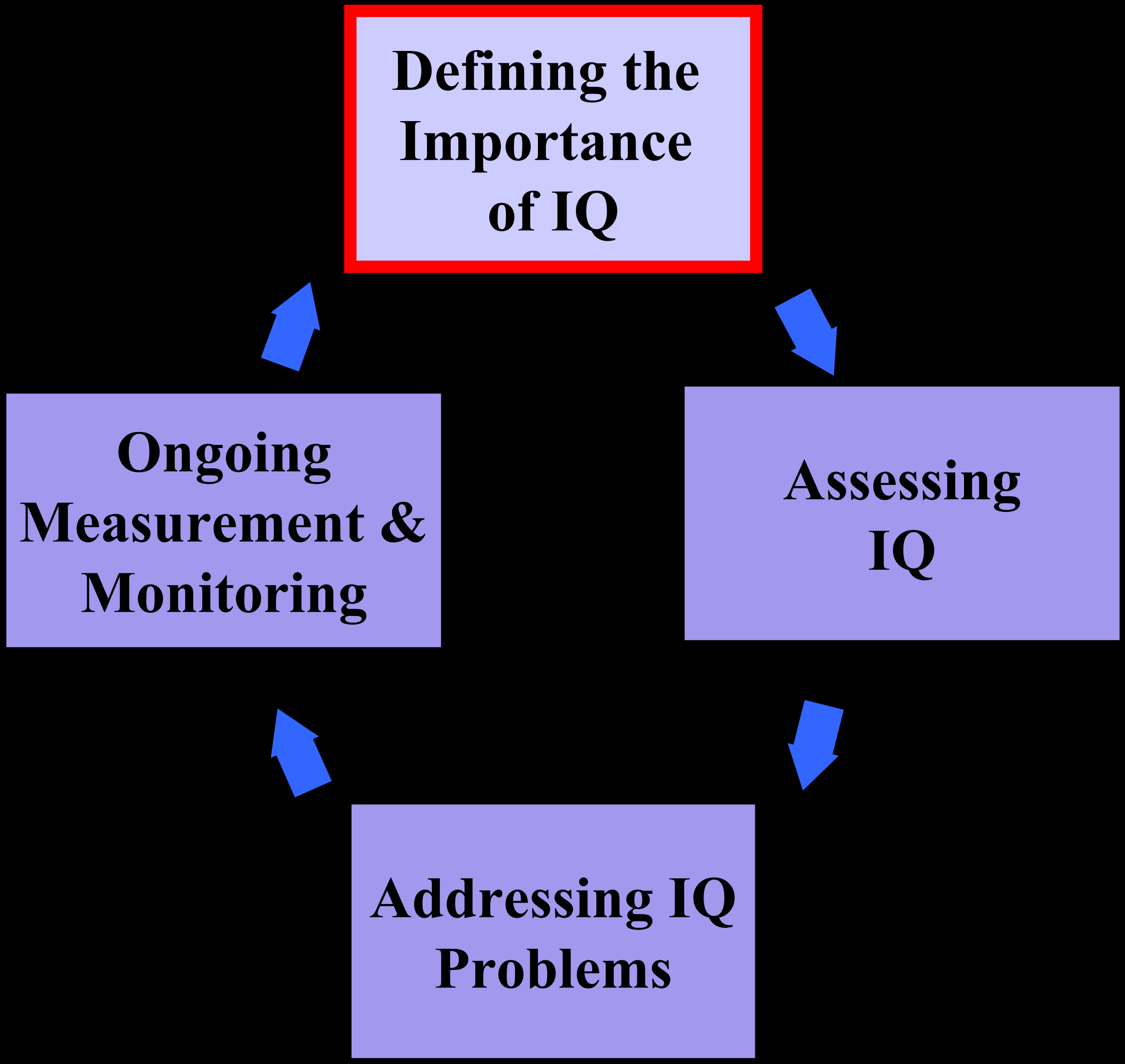
National Director,

Data Quality and Integrity

Deloitte & Touche

Deloitte & Touche Perspective on Information Quality

- Inclusion within system implementation methodologies
  - Enterprise Resource Planning (e.g., SAP, PeopleSoft)
  - Customer Relationship Management (e.g., Janna)
- Data Quality and Integrity as a part of Enterprise Risk Services
  - Data Quality Services
  - Business Intelligence Services



Defining the Importance of IQ

Assessing IQ

Addressing IQ Problems

Ongoing Measurement & Monitoring

Deloitte & Touche

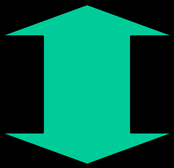
The “IQ Environment”

- IQ Environment important (English)

- Importance of the “softer side” of data quality
  - Facilitated workshops
  - Establishing an IQ task force
  - Changing the IQ environment may be political and require “change management”

The Problem of Ownership

- Information quality should be defined from the perspective of the information consumer (Wang)



- Information consumer does not control the generation (hence quality) of the information.

Deloitte
& Touche

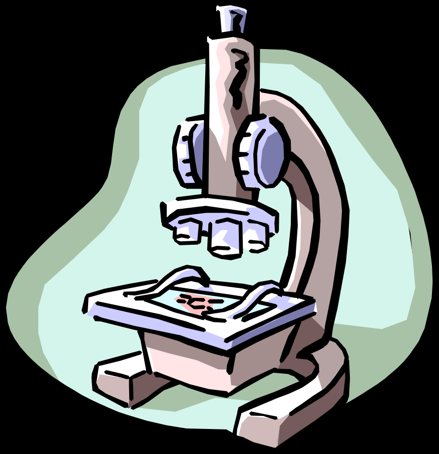
Costs vs. Benefits

- Practitioners continually need to compare the benefits of IQ to the costs of process improvement.

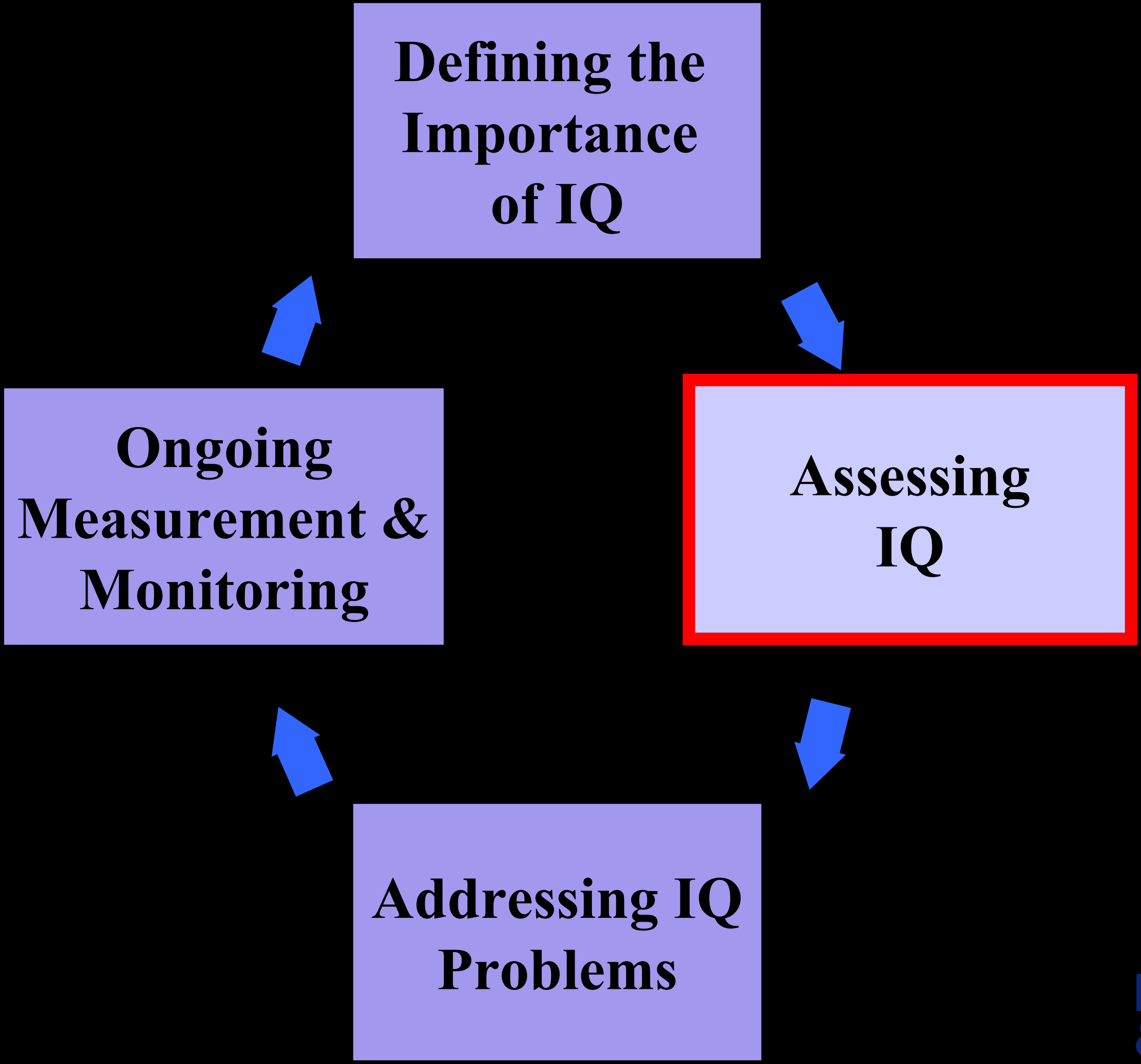
- People usually DON’T KNOW how to measure the benefits of IQ.

Research Questions

- How to measure the value of a management report?
  - What is the value of a report that is 95% accurate vs. 90% accurate? How do you obtain the measure “95% accurate” ?
  - Under what conditions is this question possible to answer?
  - How to approach the problem?



Deloitte & Touche



Defining the Importance of IQ

Assessing IQ

Addressing IQ Problems

Ongoing Measurement & Monitoring

Subjective Assessments

- Questionnaires discussed in the literature

- Benefits of facilitated workshops & interviews
  - Interview information producers and consumers
  - Weigh different priorities, perspectives
  - Subjective scoring on IQ issues can be very different from person to person

Assessing IQ

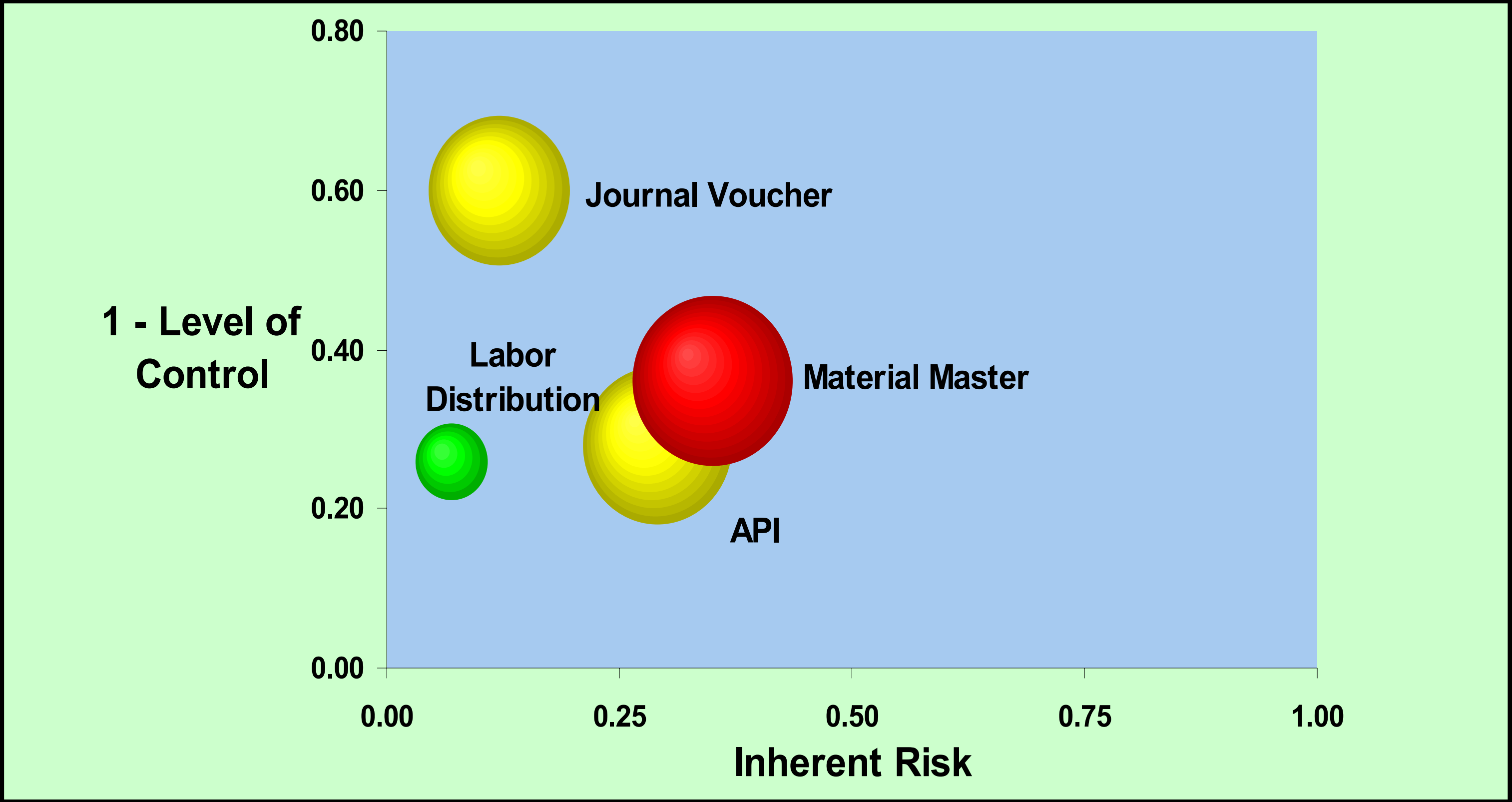
Data Analysis

Tests	Description	Examples
Base	Simple edits based on field type (individual field contents)	Numeric field must be numeric Required fields are not blank/null
Range	Business knowledge applied to an individual field (individual field content ranges) Industry norms Specific business rules	Record Code is blank, ‘08’, ‘06’ or ‘38’ Plan indicator field only contains ‘P’ Amount field has amounts >= 0 State field must contain a valid state
Intrafile	Business knowledge applied to two or more elements in the same file	Debit/credit indicator is 1 for debit, 9 for credit Cost amount is less than the Sell amount Record count field in header matches the number of records in the file
Interfile	Business knowledge applied to two or more elements in different files	Employee number is valid All customers have a Contract and Scheduling Agreement A Bill of Material Records exist for all final assembly materials in the Material Master
System / Process	Checks based on timing and completeness of data and/or system interfaces	One district only goes to one region Calculate statistics on the monetary amount field to identify anomalies

- Thorough set of tests time-consuming!

Deloitte & Touche

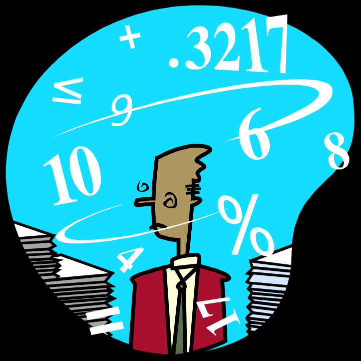
Risk Assessment



- Risk assessments can be used to prioritize work effort.

Deloitte & Touche

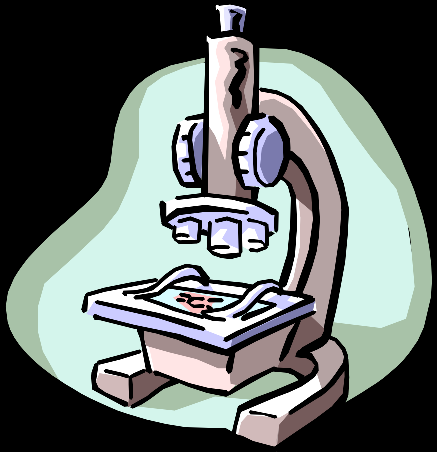
Finding Outliers

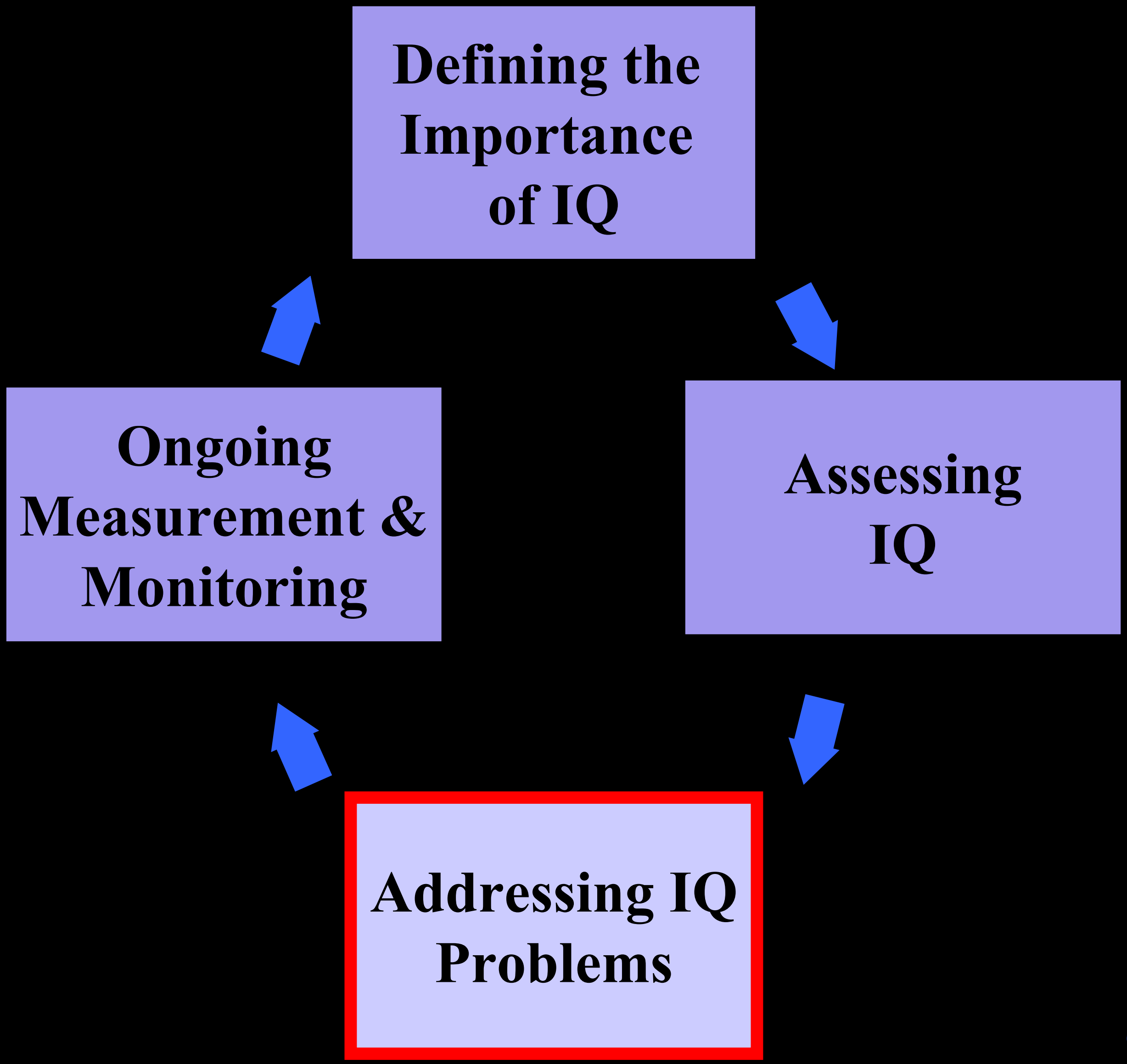


- Techniques not understood
- Advice of a data warehousing expert:
  - We will decide that today's sales total is reasonable if it falls within 3 standard deviations of the mean of the previous sales totals for that department in that store.

Research Opportunities

- Applying known methods to real-world data
  - Univariate methods
  - Other methods (e.g., Mahalanobis distances)
- Finding better methods:
  - Better ways to find outliers in categorical variables
  - Data mining in reverse? (Cluster analysis, Association rules)
  - Convex hulls?





Defining the Importance of IQ

Assessing IQ

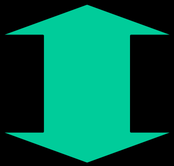
Addressing IQ Problems

Ongoing Measurement & Monitoring

Deloitte & Touche

Root Cause Analysis

- Finding and correcting problems at the source through root cause analysis is an acknowledged best practice (English, Redman).



- Reluctance, in practice, to fix problems at the source

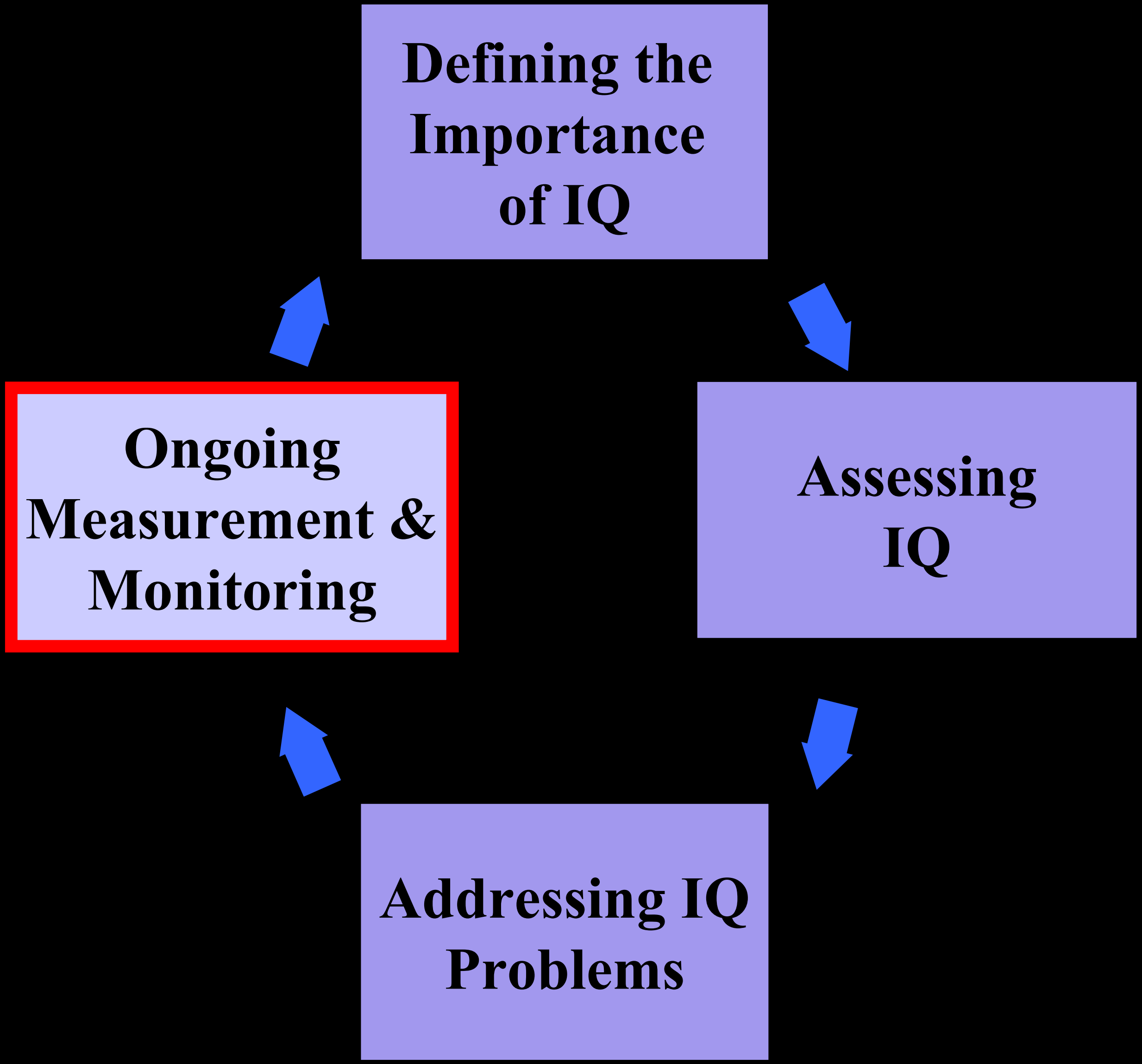
Deloitte & Touche

Research Opportunities

- Statisticians are trying to find better ways to deal with bad data (e.g., regression-based imputation).

↕

- How much effort should go into “repairing” bad data vs. demanding, facilitating, and researching better data collection?



Defining the Importance of IQ

Assessing IQ

Addressing IQ Problems

Ongoing Measurement & Monitoring

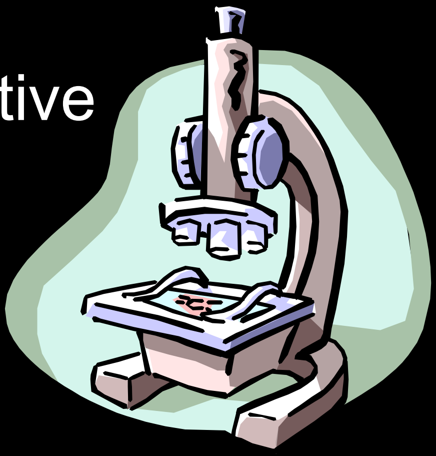
Deloitte & Touche

Obstacles

- Organizations lack summarized measurements / scores for data quality
- Without a summarized measurement, tough to prove “payoff” of root cause analysis and corrective actions
- Organizations hindered by:
  - Organizational politics
  - Lack of understanding of data quality metrics

Research Opportunities

- AGAIN: How to measure data quality?
- How to produce data quality metrics that can be summarized and monitored?
  - Technical issues of threshholds, appropriate summarization
  - May require methodologies with subjective components (like a financial statement audit)



Thank you!

References for the Practitioner

- Larry English
  - Improving Data Warehouse and Business Information Quality
- Thomas Redman
  - Data Quality for the Information Age
- Richard Wang, Kuan-Tsae Hung, Yang W. Lee
  - Quality Information and Knowledge