2019

Assessing the reliability of physical end matching and chemical comparison of pressure sensitive tapes

Meghan Nicole Prusinowski
mnp0006@mix.wvu.edu

WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2019

Assessing the reliability of physical end matching and chemical comparison of pressure sensitive tapes

Meghan Nicole Prusinowski

Assessing the reliability of physical end matching and chemical comparison
of pressure sensitive tapes

Meghan Prusinowski

Thesis submitted
to the Eberly College of Arts and Science
at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Science in
Forensic and Investigative Science

Tatiana Trejos, Ph.D., Chair
Jacqueline A. Speir, Ph.D.
Andria Mehltretter, M.S

Department of Forensic and Investigative Science

Morgantown, West Virginia
2019

Keywords: Duct tape, electrical tape, error rates, physical end matching, similarity scores,
X-ray fluorescence, elemental analysis, trace evidence, forensic science

ABSTRACT

Assessing the reliability of physical end matching and chemical comparison of pressure sensitive tapes

Meghan Prusinowski

Pressure sensitive tapes are a common evidence type in a variety of forensic cases, including violent crimes, drug trafficking, and terrorism. Forensic laboratories are often requested to analyze tape samples for composition and potential source identification. Tape samples manufactured from different sources are often distinguishable when using a combination of sensitive analytical techniques. Nonetheless, the interpretation of the data remains challenging due to a lack of standard criteria and guidelines from which examiners can support their conclusions. As a result, there is a need to standardize the protocols of interpretation of physical and chemical comparisons of tape evidence.

The manufacture and composition of duct tapes and electrical tapes contribute to physical and chemical properties that require different approaches for source attribution. Duct tape physical features are highly variable and can be used to distinguish between different samples. For samples with consistent physical characteristics, the distinctive realignment of their edges indicates the samples were once part of the same piece. However, regardless of the evidential value, there are no defined protocols on how to substantiate the identification of a fracture fit. Conversely, in electrical tapes which have fewer physical features to evaluate, the elemental analysis is often more informative than the physical examination.

The long-term goal of this study is to assess the reliability of physical end matching of duct tape and chemical comparison of electrical tape to provide more objective criteria to inform the examiner's opinion. The first goal of this study was to develop a systematic protocol to quantify the quality of a fit between duct tape edges, assess the performance of the method, and provide a statistical assessment of the weight of the evidence. The quantitative criteria proposed is an edge similarity score (ESS) - the relative number of matching scrim bins across the tear out of the total number of bins along the length of the fracture. The scrim bins provide a consistent and quantitative measure of the quality of a match and means for a systematic peer review process. A set of 2280 duct tape end comparisons (including hand-torn and scissor-cut edges, and stretched and pristine samples of different quality grade) were evaluated to validate the method. The evidentiary value of physical fits was evaluated through similarity metrics, probabilistic estimates from the distribution of ESS, and score-based likelihood ratios (SLR) to offer statistical support for a match/non-match conclusion. No false positives were found for any of the sets. The accuracy for the low and mid-quality sets ranged between 99.5-99.8% with false negative rates of 1-2%. The high-quality set showed higher uncertainty, with a 21.4% false negative rate, and overall accuracy of 84.9%. On average, ESS higher than 80% provided a score-based likelihood ratio (SLR) that supported the conclusion of a match, while ESS lower than 25% provided an SLR supporting the conclusion of non-match.

The second objective was to provide enhanced reliability of chemical comparisons of electrical tapes. X-ray fluorescence (XRF) was proposed as an alternative elemental analysis technique for

the chemical comparisons of electrical tapes. While elemental analysis by scanning electron microscopy-energy dispersive spectroscopy (SEM-EDS) is currently the most informing method for the characterization of electrical tape backings, XRF offered superior sensitivity and enhanced discrimination capabilities. A set of 40 electrical tape backings known to originate from different sources was used to assess discrimination and classification capabilities of the method. The discrimination for this tape set for XRF ranged from 81.5-91.0% depending on the instrumental configuration, whereas SEM-EDS only achieved a discrimination power of 78.8%. XRF also showed to be comparable to the discrimination capability of LA-ICP-MS (84.6% discrimination power) for this set of tapes. In addition, a semi-quantitative method for estimation of significant differences among samples was evaluated to complement spectral overlay.

This study developed a quantitative evaluation of the quality of a fracture match and offered an empirical basis that substantiates the reliability of duct tape physical fits and offers statistical support for examiners to inform their opinion. Moreover, it is anticipated that XRF will be easily incorporated in the near future for electrical tape analysis, since it is a technique that is widely applied in crime laboratories for other materials and shown to offer superior performance than conventional methods. The methodologies applied in these studies align with current protocols and instrumentation available in forensic laboratories and lay a foundation for the future development of additional systematic methods for other trace materials.

Acknowledgements

TABLE OF CONTENTS

1. Introduction

1.1 Tape Structure and Manufacturing

The term pressure-sensitive tape refers to tapes that are made with a film or cloth backing, as well as an adhesive layer that adheres to a surface upon application of light pressure. Figure 1.1 illustrates the primary anatomy of tapes. The most common components of tape are the release coating, the backing, and the adhesive layer. A small layer of primer is applied between the backing and adhesive but not often detectable during forensic examination.
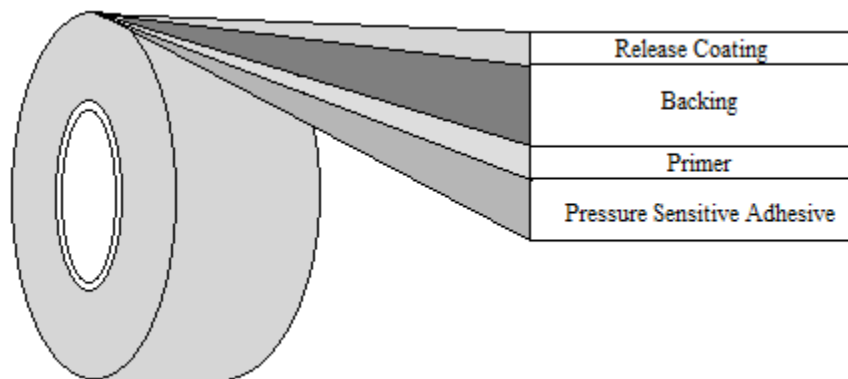
Figure 1.1. A diagram showing the four most common layers in pressure sensitive tapes (not to scale).

There are various kinds of pressure sensitive tapes on the market, and some contain additional layers that contribute to the end use of the product. Table 1.1 lists the main type of tapes available in the market and their primary end-use.

Tapes are used in a wide variety of criminal activities, including violent crimes, drug packaging, and improvised explosives. In 2012, the Scientific Working Group for Materials Analysis (SWGMAT) conducted a survey, representing responses from 105 crime laboratories from 18 different countries. From the participant labs, 92% of them had a trace evidence unit, and from those 95% performed tape analysis. Duct tape is the most prevalent tape evidence found in casework in the US and abroad, followed by electrical and packaging tape (Figure 1.2) [1].

Table 1. 1. Main types of pressure-sensitive tapes available for consumers and their main uses.

Type of Tape	Main Anatomy	Chemical Composition	End Use	Example Image
Duct Tape	Polymer Backing Fabric Reinforcement Adhesive	Polyethylene film Cotton/polyester fibers acrylic, rubber or silicone adhesive	Waterproof tape used for general application to a wide variety of materials indoors and outdoors	
Electrical Tape	Polymer Backing Primer Adhesive	Polyvinyl chloride film Latex acrylic, rubber or silicone adhesive	Wrapping and insulating electric wires	
Packaging Tape	Polymer Backing Adhesive	Polypropylene or polyester film acrylic, rubber or silicone adhesive	Sealing packages	
Strapping Tape	Polymer Backing Glass Reinforcement Adhesive	Polypropylene or polyester film Fiberglass Rubber	General application where extra reinforcement is necessary	
Gaffer's Tape	Fabric Backing Adhesive	Cotton backing film Rubber	Cover wires without damaging surfaces when removed	
Office Tape	Polymer Backing Adhesive	Cellophane, cellulose, or polypropylene film Acrylic or rubber	Adhere materials such as paper	
Foil Tape	Metal Foil Backing Adhesive	Aluminum, copper, lead or steel foil acrylic, rubber or silicone adhesive	Heat resistant tape usually used to seal heating/cooling ducts	
Medical Tape	Fabric Backing Adhesive	Fabric or paper backing Hypoallergenic adhesive	Adhere bandaging/gauze to skin	
Friction Tape	Polymer/Cloth Backing Adhesive	Cotton or synthetic fiber cloth with rubber adhesive imbedded	Wrap around materials to make them easier to handle	
Masking Tape	Paper Backing Adhesive	Paper backing Rubber or acrylic adhesive	Used when painting to cover areas that should not be painted but not damage surfaces on removal	
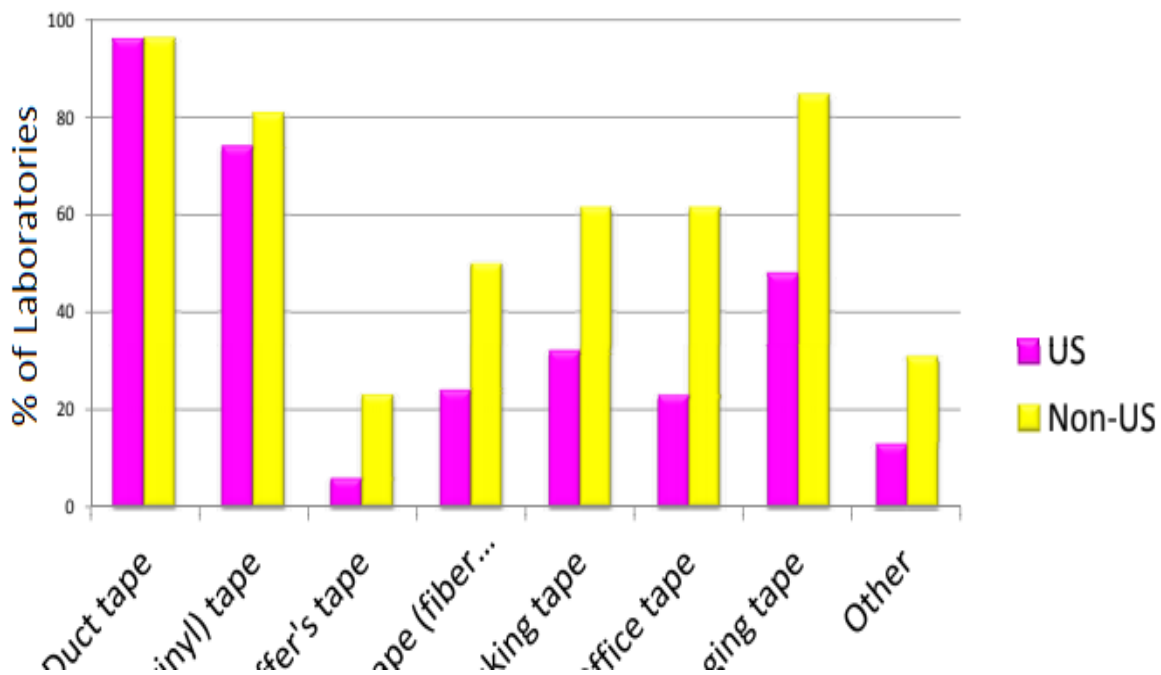
Figure 1.2. Distribution of type of tapes most commonly received as evidence in trace evidence laboratories. (Others include paper backed tape, medical tape, painters' tape and foam tape) [1].

The manufacturing process for tapes introduces characteristics that are reproducible within a single roll but can be used to distinguish between samples originating from different manufacturers. The liquid adhesive used in tapes is formed from a combination of rubber (natural or synthetic), extenders, and tackifiers. The adhesive is rolled onto the backing layer through a process known as calendering. The backing layer is primarily polymer-based, with duct tapes mostly containing the polymer polyethylene, whereas electrical tapes usually containing polyvinyl chloride, and more rarely polyester or polyimide. In addition to the backing and adhesive, duct tape contains a fabric reinforcement layer known as the scrim that contributes extra strength to the tape [2-5].

During the manufacturing process, the drying tapes are rolled into large rolls known as jumbo rolls. Jumbo rolls are cut to size for the width of tape to be sold to customers. The manner in which tape rolls are distributed can complicate efforts to determine the original source in a forensic investigation. The companies that manufacture the tape rolls may supply cut rolls from a single jumbo roll to multiple distributors. From there, the distributors may label tapes with their own

brand name before selling the tape. The result is that tape rolls purchased at two different supply shops may have similar physical and chemical properties.

A diagram of the typical anatomy of a duct tape and electrical tape is shown in Figure 1.3. The different compositions of tape backings and adhesives contribute significantly to the end function of the tape. In duct tapes, the backing color and texture, along with the pattern and type of fibers used in the scrim vary by manufacturer and tape function, and so act as major distinguishing characteristics between sources.
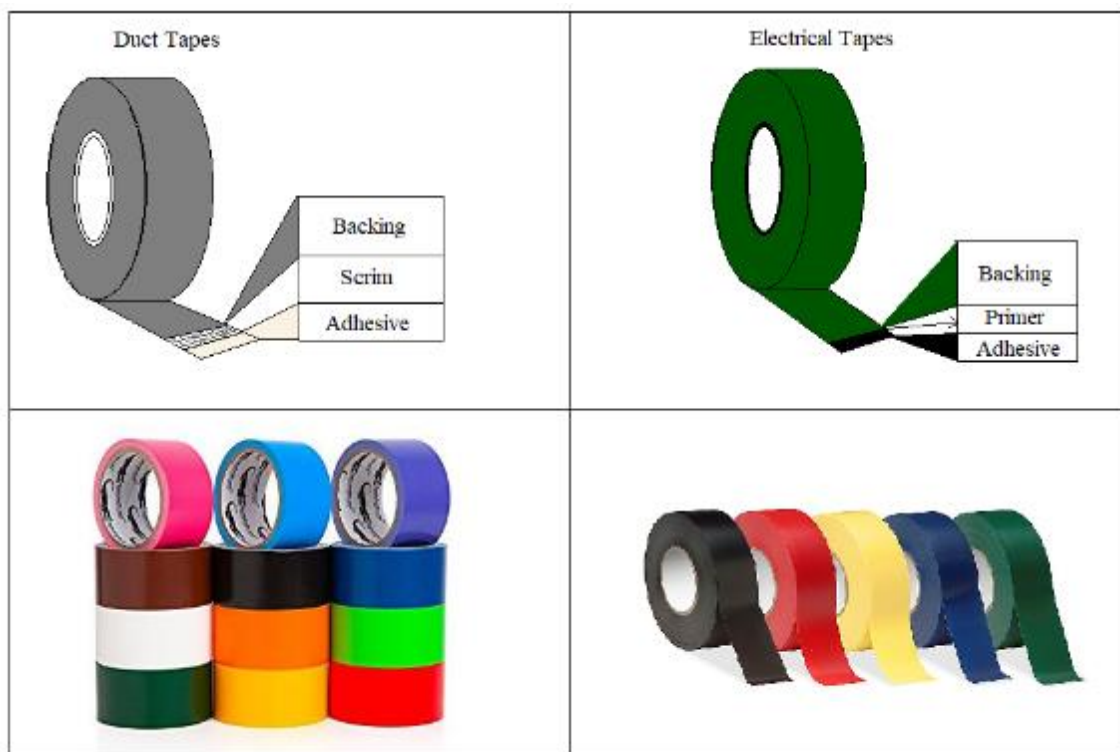
Figure 1.3. Diagram showing the main layers in duct tapes and electrical tapes (not to scale).

Certain characteristics, such as the width and thickness are more likely to be altered if the tape has been stretched or subjected to weathering. Other features are more subjective to the opinion of the examiner, such as the adhesive color. Due to the variety of features that can be observed, physical comparison can distinguish many tapes prior to any chemical classification. Figure 1.4 illustrates some distinctive physical features observed in duct tapes. Duct tapes can come in many colors and patterns, but the most prevalent ones are gray and black backing.
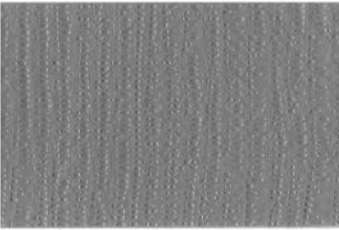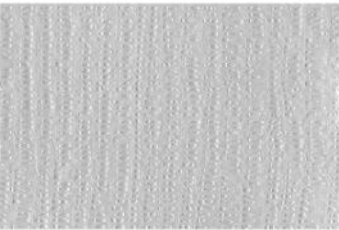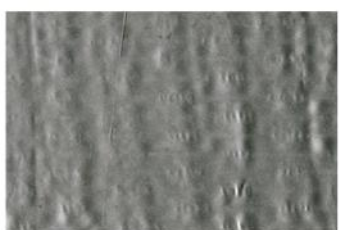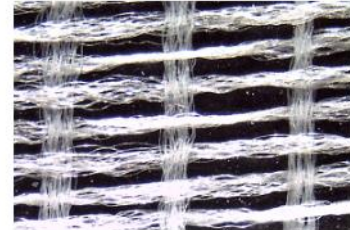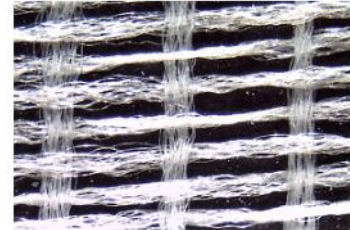
Figure 1.4. Examples of main physical features observed in duct tapes.

In duct tape comparisons, the physical features have shown to provide a significant discrimination potential. For instance, Mehltretter and Bradley examined 82 different duct tape samples collected over a period of time ranging from 1993 to 2003 and observed the physical characteristics and FTIR spectra of the backings and adhesives. The authors concluded that 99.6% of the duct tape sample set was differentiated with just the physical features [4]. Overall, the visual examination was sufficient to distinguish the majority of the tape samples.

Since tapes are man-made mass-produced items, single rolls are not necessarily different from others produced from the same company. Many rolls can be cut from the same jumbo roll, and so can share similar physical characteristics. Studies have been conducted to evaluate the variability of tape features throughout the manufacturing jumbo roll, as well as the variability of the physical characteristics within a single roll of tape. As found by LaPorte and Wiemer, the width and scrim count of the tapes were consistent throughout a single roll of tape. Nonetheless, the authors reported that the backing thickness and adhesive color varied considerably throughout the entire roll, so those characteristics alone should not be used when examining duct tape samples [5]. In addition, Mehltretter et. al found that rolls cut from the same jumbo roll had indistinguishable physical and chemical compositions [6]. The significance of the low variability throughout the rolls lends support, in the event of a source association, that the tape either originated from the known source, or from another roll manufactured in the same manner. Due to the large diversity of tape products, a source association can provide valuable evidence in trials and investigations. However, as with any mass-produced materials an "association" conclusion between tape samples in a forensic analysis may be limited to the class characteristics of the samples. Individualizing features in tape comparisons may only arise from a physical match between tape ends [2,3].

Unlike duct tapes, electrical tapes do not have a scrim layer. The lack of the scrim layer makes electrical tape less strong, but more flexible. The composition of electrical tape also keeps the tape from conducting electricity. Contrary to duct tapes, which can have an extensive variety of colors, textures, and patterns on the backing and fabrics, electrical tapes are more likely to be limited in the types of patterns and colors most commonly used by consumers. The limited colors and higher degree of flexibility can make electrical tape samples more difficult to distinguish at a physical examination, and physical fit conclusions are more challenging due to the stretching of the fracture edges.

Conversely, the chemical profile of electrical tapes can be used to distinguish between sources manufactured with different raw materials. For instance, the main polymer and plasticizers to increase flexibility can be identified through various analytical techniques. Other components contribute to the composition of the backing, like heat resistant additives containing inorganic elements Na, Mg, Sr and Nb, polymer catalysts using Li and B, metallic dryers containing organometallic compounds such as Cu, Zr, Sr, Pb, Zn, and Bi, and flame retardants such as salts of Br, Mo and Sb [7-13]. The elemental profile serves as a major distinguishing factor between electrical tapes originating from different sources [12-13].

Table 1.2 summarizes the main chemical composition of duct tapes and electrical tapes. Chemical analysis is often conducted by instrumental methods that elucidate major, minor and trace organic and inorganic constituents as described in section 1.2.

Table 1.2. Main chemical composition of duct and electrical tapes.

Tape Type	Layer	Main Components	
Electrical Tape	Backing	Polyvinyl Chloride (most common), Plasticizers, Stabilizers, Flame retardants, Processing Aids	
	Primer	Water, Latex	
	Adhesive	Rubber – natural or synthetic	Styrene/butadiene Styrene-Isoprene-Styrene (SIS) Butadiene-Isoprene-Styrene (BIS) Polyisoprene
		Fillers	Clay, calcium carbonate, talc
		Tackifiers	C5, C9, terpenes, rosin derivatives
		Curing Solvent, Solvent Blend	
Duct Tape	Backing	Polyethylene (most common), Pigments	
	Scrim	Cotton, nylon, polyester, rayon	
	Adhesive	Rubber – natural or synthetic	Styrene/butadiene Styrene-Isoprene-Styrene (SIS) Butadiene-Isoprene-Styrene (BIS) Polyisoprene
		Fillers	Clay, calcium carbonate, talc
		Tackifiers	C5, C9, terpenes, rosin derivatives
		Antioxidants	Amines, hydroquinones, phenols
		Oils	Naphthenic, polybutenes
		Pigments	Titanium dioxide, zinc oxide, carbon black
		Crosslinker	Sulfur, phenolic & quinoid resins, peroxides

1.2 Forensic Analysis of Tape Samples

When forensic tape samples are processed as trace evidence, there are typically two main goals for the analysis. If possible, the examiner will attempt to characterize the composition of the tape and compare it to available manufacturer product information to determine the source of the tape. However, maintaining reference databases and sources of manufacturing information for tape samples is challenging due to the way tapes are distributed [2]. This type of information is particularly useful during investigative stages where no control comparison sample are available (i.e., no suspect tape roll found yet). While there are reference databases for tapes, such as the National Forensic Tape File, the widespread variability of products makes the creation and upkeep of the databases difficult, and as a result there is not widespread use of them in most laboratories.

On the other hand, in cases where both questioned and known tape samples are available, they will be compared to identify any significant differences or commonalities. By evaluating tape samples for physical features and composition, known and questioned samples can be associated with class characteristics if there are no significant differences identified throughout the analysis. Significant differences between samples include any distinct physical or chemical variance that would allow the examiner to infer the samples originated from different sources. There are various kinds of analytical techniques that can be performed on tapes. Many provide complementary information, and so the choice of what techniques should be performed relies on each laboratory's available instrumentation, and the samples in question. Methods for the analysis of tape include microscopy (stereomicroscopy and polarized light microscopy (PLM)), Fourier Transform Infrared Spectroscopy (FTIR), Scanning Electron Microscopy-Energy Dispersive Spectroscopy (SEM-EDS), X-ray Diffraction (XRD), and Pyrolysis Gas Chromatography Mass Spectrometry (Py-GC/MS) [16]. More recently, methods such as laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS), laser induced breakdown spectroscopy (LIBS), and X-ray fluorescence spectrometry (XRF) have been explored for tape comparisons [2,14-18].

Typically, examiners will start analysis with a physical and microscopical examination of the tape samples, as the method is fast and non-destructive, and allows for rapid differentiation between products originating from different sources. Duct tapes in particular are highly variable in appearance, with characteristics such as the backing color and texture, adhesive color, thickness,

width and the number and pattern of the scrim layer all used to distinguish between different products [2]. A study in 2012 evaluated 82 separate duct tape rolls to determine how many could be differentiated using common analytical techniques used for tape. Only by physical characteristics, 99.6% of the duct tapes were distinguished. Using the data collected from the rest of the analytical techniques increased the discrimination power to 99.8% [4]. Electrical tapes can also be discriminated at the physical level by looking at the color, thickness, finish, and surface characteristics of the samples. However, physical features for electrical tapes are not as variable as in duct tapes, and so do not offer the same power of discrimination. Mehltretter et. al in 2011 found that of 90 black electrical tape backings examined, the physical examination only resulted in a discrimination power of 64.3%. However, when FTIR, SEM-EDS and Py-GC-MS were added to the analytical examination, the discrimination improved to 94.3% [12].

When no significant differences are found in the physical features, the chemical composition is identified using a series of techniques. At each step the composition of the questioned samples will be compared to that of the known in order to identify major components and determine if the samples originated from different sources. Common practice is to first use techniques that are minimally destructive, to ensure that enough sample remains for any future re-analysis. Fortunately, forensic tape samples are more likely to be recovered in larger amounts than other kinds of trace evidence. After all the recommended analysis techniques are exhausted, and if there are no significant differences found, then an association based on class characteristics between the known and the questioned samples will be reported. The significance of the association depends on the rarity of the observed features in a relevant population and the discrimination potential of the methods used during the examination.

Studies have shown that the variability of the physical features and chemical composition within a roll of tape is much lower than the variability between samples [2-6,12-18]. However, due to the complex process of mass-manufacturing and distribution of tape to supply shops, a tape cannot be associated to a single source. The exception occurs when a physical fit between two tape ends provides individualizing characteristics that associate the two tapes to one another exclusively. Class-associations, nonetheless, are still very valuable pieces of information that can help during investigation, reconstruction of events, establishing important links among cases, or as probative evidence in trials.

9

1.3 Physical Fit Evaluations in Forensic Science

Physical fit determination, known formerly as fracture matches, is the process of re-aligning two items in an attempt to determine if they were once part of the same object. The use of physical fits in forensic science has been ongoing for decades for a variety of materials, including but not limited to paper matches, glass, paint, tape, and plastic. The identification of a fit between edges is an important attribute when comparing samples, as physical fits are the highest level of association for trace materials because the conclusion shifts from associating samples based on class characteristics to the identification of a single source. To date, the major assumption when comparing materials for physical fits is that a "random" physical fit between materials that did not separate from the same source is unlikely and serves as the foundation for its evidential value.

Glass

Glass is difficult to observe for fracture fits due to the reflective properties, making illumination critical to the observation of fracture features. In 1959, a study was conducted by Donald Nelson to evaluate glass fractures to determine what features of glass would assist in the determination of a fracture fit. He found that hackle marks, lines formed in the glass during the fracture, are visible in the glass under correct illumination, and when photographed in alignment and slightly displaced offer the most information towards the alignment of glass fragments [19]. In 2016, Baca et. al. published a study in which they evaluated the breaking patterns of 60 panes of glass, 60 clear wine bottles, and 60 polymeric tail light lenses. While they were not specifically attempting to evaluate physical fits, they did observe some characteristics of the fracture patterns that are critical to the evaluation of a physical fit between edges. Namely, they found that the velocity of the breaking force, and whether the breaking object was sharp or blunt influenced the overall appearance of the fracture. They found that using a blunt tool or object leads to more radial fractures, and that static objects will have more complex patterns than dynamic fractures [20]. Fractography research that use systematic breaking or separation methods to understand the factors that influence the formation of the fractures and the prediction of sources of damage have provided a valuable body of knowledge [21]. However, no studies have directly evaluated the accuracy of glass fragment physical fits.

Paper matches

Physical fits have also been evaluated in paper matches. Paper matches are one of the most inexpensive kinds of matches available on the market. Usually packaged in small books, the matches are typically made of a stiff paper or cardboard material stapled into the books. There have been a few studies evaluating whether or not separated matches could be associated back with their original book. A study in 1967 by H.J Funk evaluated if 50 books of paper matches could be traced back to their original book just using the physical characteristics. Funk found that the matches typically did not separate from the book with a clean edge, and that extraneous fibers could be found extending down from the match and from the book. The fibers in undyed matches could also be contrasted by staining them with phloroglucinol in hydrochloric acid. Using the fibers and the separated edge features allowed Funk to determine which matches were separated from a given book [22].

Another study in 1986 evaluated 41 match books using two lasers, an argon laser (457-514nm) and a dye laser (593nm) with different barrier filters with a magnifying glass and a stereomicroscope. The lasers were used in order to highlight luminescing inclusions to the cardboard and matches, as well as any fluorescent fibers. The author found that using the argon laser showed that 59 matches out of 120 (around 15 books) had two or more luminescing inclusions that corresponded with inclusions on adjacent match. The dye laser resulted in 17 books with two or more luminescing inclusions and a total of 35 matches out of 120 had two or more luminescing inclusions corresponding with the adjacent match. The authors noted that some inclusions could only be seen with one laser or the other, such as the surface fibers which were not visible with argon laser under the magnifying glass [23]. The application of physical fits for paper matches shows how the application of alternative light can highlight additional features that aid in the comparison. The methods used to identify the matches to their source were based on the qualitative features of the match and separated book, and did not utilize a quantitative approach.

Tsach et. al. demonstrated that the torn patterns of different materials remain unique, even under reproducible conditions of the tearing process [24]. An important assumption that can be made with paper materials, which does not apply to other mass-produced items, is that the microfibers tend to be randomly deposited during the manufacturing process. Since the fibers are usually randomly oriented, their alignment across torn edges is considered to provide unique features. The

assumption has been experimentally observed in casework, but its scientific reliability has not been formally proved.

Paint

Paint flakes can often be recovered from crime scenes. Automobile paints in particular hold great value when conducting a physical fit due to their 3D-multiple layer structure. While it is not as common to get large enough intact paint fragments to regularly perform physical fit evaluations, there have been cases where paint flakes have been used to establish an association. A case report published in 2001 involved a stolen safe. Paint fragments found at the home of a suspect were examined and compared to the paint from the door of the safe of interest. During the examination, six weld beads on the safe door were found to be missing paint, and so casts were made of the beads and images were taken of casts and paint flake backs for comparison of ridges. When compared, the patterns on the back of the paint fragments and the patterns on the weld beads were found to be consistent. The authors claimed that the welding ridges could be considered unique due to the high variability of pattern formation in the welding process due to manual action of welder along with external factors such as the ambient temperature, the metals used, the speed of the process, and the type of weld [25].

Plastic

Consecutively torn plastic bags, such as garbage bags, have also been shown to have features that allow for physical fit determinations. Polarized light can be used to reveal various details in thin plastic, such as additives that appear as patterns or striations. The striations can be compared across the separated edge to show which bags were sequentially removed from the roll. Horizontal dark streaks in the material formed during the manufacturing process are classified as features which can help individualize the bags. The classification of pigments are fisheyes, which manifest as randomly-distributed dark pigments, arrowheads which are seen as dark triangular striae, horizontal striae of dark pigment called tiger stripes, and straight lines called die lines [26]. A study of plastic bags in 1983 examined both garbage bags and sandwich bags. The garbage bags included ten packages of different brands of garbage bags from local stores, 13 consecutively-made garbage bags from a plant, and three retail packages of consecutively-made garbage bags from two different plants. The sandwich bags included five packages of different brands from local stores, and two packages of consecutively-made sandwich bags from a plant. The bags were

observed to note the color, perforations, construction, code, pigment bands, and presence or absence of hairline marks. For garbage bags, production sequence was determined by finding the slope of a prominent marking across all bags. They were then examined for colored striations under polarized light, for individual characteristics including fisheyes, arrowheads, streaks, and tiger stripes. Individual characteristics examined on sandwich bags included the surface scratches and colored bands. The authors concluded that the bags could be found to be consecutive when both class and individual characteristics aligned. They also noted that understanding the manufacturing process can be utilized to make inferences about the order of markings across multiple plastic bags [27].

An additional study noted that surface patterns on plastic could be seen in better detail by lightly dusting the surface with fingerprint powder, and that the additives to the plastic material is abrasive and causes machine part wear, leading to differences in perforations, cut edges, and roller imprints that are then visible in the sequentially removed bags [28].

Other case reports

Physical fits have also been used in more unusual cases as a means of connecting a suspect to a scene or reconstructing the events. A case study in 1963 involved the comparison of a broken part of a fingernail found at the crime scene to fingernails from suspects. The examiners used the microscopic striae underneath the nail plate as the foundation for the comparison, and identified one of the suspects as a match. They claimed that the striae under the nail growth are similar to fingerprints and are distinct to each person, although that claim remains unsupported [29].

Another case published in 1975 used ultraviolet light to visualize patterns in adhesive left on the separated heel of a shoe found at a crime scene to the shoe of a suspect. The examiners did have some markings from the nails in the heel, but the nails alone were not enough to finalize a conclusion. Using the ultraviolet light caused the adhesive on the heel and the other part of the shoe to fluorescence. The examiners photographed the shoe and heel under the ultraviolet light, used the negatives to compare the patterns, and found the shoe to match the heel [30].

A case study in 1984 reported a physical match found between a length of rope found at the crime scene and a spool found in the suspect's home. The rope was examined to determine the diameter,

direction of twist, number of twists per unit length, material used to construct the rope, number of strands, threads and fibers. The particular rope in the case contained two orange fiberglass cords, one of which had a distinct fracture that matched the other end from the spool [31].

Another case in 1985 connected a suspect to the crime scene using a portion of skin left behind. The section of skin was quite large at 1.25 inches long, and contained several layers of skin. The skin was preserved so as to maintain the moisture and prevent shrinkage. It was also noted that the skin had ridge detail, indicating that it was off of a finger or toe. Two days later, a suspect was found with a deep cut in his thumb. When the skin was compared to the injury on his hand, it was found that there was a match, not only through the shape and size of the missing skin, but also through the ridge detail that was consistent across the edges. The authors of this case did note that the match was possible in this case because there was a large enough section of skin left behind to do a physical comparison, and that time between the injury and the comparison had been short enough that the injury had not healed much nor allow the skin fragment to degrade [32].

Although case reports illustrate the potential relevance of physical fits, it also reveals the lack of scientific support that examiners can use to inform their opinions and assess the significance of their findings. To date, few studies have addressed performance measures such as error rates [22,33-38], and fewer have utilized a blind process or a significant sample size [35-38].

Tape

A SWGMAT survey conducted in 2012 indicates that physical fit examinations of tapes are common. 97% of the participant forensic laboratories reported they performed physical match examinations on tape ends. Physical fit examinations of tapes were conducted primarily by trace examiners (82%), toolmark examiners (8.5%) and chemist examiners (6.5%) [1]. Consequently, there have been recent efforts to address the scientific reliability of fracture fits, particularly for electrical tapes and duct tapes.

A case study in 1986 looked at physical fits between electrical tapes in the context of a case where a hand grenade caused an explosion during a demonstration. The hand grenade had been originally wrapped in electrical tape, of which a small section was collected and found to be a physical match to the tape wrapped around another grenade in the possession of the explosive supplier [39].

Weimar reported different ways to improve the identification a physical match in electrical tapes, which are more prone to distortion and stretching than other materials. His study in 2008 evaluated how heating the sample using hot air for a few seconds could cause stretched samples to return to their original state, and therefore allow the physical fit to be identified [40]. He followed with another study in 2010 evaluating scissor-cut electrical tapes. He devised a method to place the samples in hot distilled water to reduce stretching, and then preparing silicone casts of the edge to compare samples. The author observed that features on the surface of the scissors were imparted into the edge of the tapes across the tear for matching samples [41].

A study evaluating the performance of comparing electrical tape edges was published in 2011 by Bradley et. al., where out of 106 true matches in the study, 8 were not identified, and 1 edge was reported as a false match. The number of true non-matching edges in the study was 2036, resulting in a false positive rate of 0.049%. The authors concluded that with electrical tapes, the potential for stretching and distortion can cause examiners to be more conservative in their approach to physical fits, or potentially result in association between samples of different origin [35].

A similar study was conducted in 2006 that examined duct tapes instead of electrical tapes. With four examiners participating, 92% of the total 81 hand-torn and scissor-cut tape ends were correctly identified, and no false matches were reported. The authors also noted in this study the importance of peer review in physical fit determinations, as some of the tapes that were falsely excluded were determined to be matches through a blind peer-review process [36].

McCabe et. al. in 2013 conducted a comprehensive study of duct tapes. Eight different duct tapes were used with one of four separation methods, including hand-torn, scissor-cut, box knife, or Elmendorf tear tester for a total of 2400 tape pairs. Overall, the study found accuracy for torn tape to be 98.58-100%, and for cut tapes to be 98.15-99.83%, with false positive rates under 3.33%, and false negative rates under 2.67%. This study was the first to date to assign a quantitative value to the edge of a tape in addition to the overall match/non-match conclusion. The participants reported a score measured by the relative length of the fracture with matching features respective to the entire width of the sample. However, despite collecting the scores for each tape, no other statistical analysis was reported to assess the weight of the evidence [37,38].

In recent years, the National Academy of Sciences (NAS) report in 2009 and the 2016 President's Council of Advisors on Science and Technology (PCAST) report investigated the areas of forensic science that needed a more rigorous scientific foundation and a clear establishment of limitations. Within those reports, pattern evidence was an area of major concern, as there were many techniques in the discipline that did not have enough scientific substance to support the claims that were made in the courtroom. According to the reports, some analysts in these disciplines were liable to overstate the weight of the analysis and the evidence in their testimony [42,43]. Since the publication of the reports, there have been many efforts to address the concerns surrounding pattern evidence. The development of international standards and the publication of many peer-reviewed studies examining the basic assumptions and foundations for methods in the field have contributed to more objective approaches to the interpretation of the results. In that regard however, physical fits have not been an area of much focus in recent years, until in 2018 the NIST-OSAC subcommittee highlighted the need for more objective and quantitative approaches to physical fits in trace evidence materials [44]. There have been a few studies that evaluated the overall error rates for examiners performing physical fit determinations, which found that examiners are able to demonstrate relatively high accuracy, with low error rates when making large numbers of comparisons [35-38]. There have also been a few studies that evaluated automatic algorithms for the comparison of edges, but the outcomes of those algorithms have still not been as successful as what a human examiner is capable of [45,46]. One such algorithm created by Yekutieli et. al. assessed tears of three kinds of materials: paper, silicone and Perspex (similar to Plexiglas). The total error rates (accounting for both false positives and false negatives) was 0.007 for silicone, 0.37 for paper, and 0.4 for Perspex. The authors noted that the results were lower than expected, but that they only used fracture lines, and not additional 3D features that experts would consider. As such, they suggested adding surface details would bring their algorithm closer to the results of experts [45]. Another algorithm designed by Ristenpart et. al. to examine duct tape edges assigned a sum of square residuals (SSR) value to each tape edge. Low SSR values were indicative of true matches, while high SSR indicated non-matches. While the authors found that the algorithm was able to assign the lowest SSR to the matching edge in 97% of the tears, the algorithm also had problems with false positives, especially in scissor cut samples, in which the false positive rate was 61% [46]. A major goal of this research was to develop a systematic, quantitative method to describe the quality of a tape edge match, and to use statistical methods to interpret the data. The

systematic method for tape developed in this study will serve as a foundation for the application to other trace materials.

1.4 Statistical Analysis of Physical Fit Determinations

The statistical evaluations conducted in this research aim to provide empirical evidence of the significance of duct tape fracture fit conclusions. Since the physical features left after separation of a fractured tape edge cannot be reproduced—even under controlled tearing conditions—the assessment of the reliability of a physical fit is based on the interpretation of qualitative and quantitative data derived from the comparison of edge ends of "known true origin" [47]. In other words, the performance of the method can be measured by creating a validation dataset composed of samples known to have been separated from the same object (true matches) and samples known to have been separated from different objects (true non-matches). The purpose of the validation set is to evaluate empirically how often the comparison method leads to misleading results.

A validation dataset should be able to simulate the samples as typically received in a forensic case as close as possible. In our case, this would be, for instance:

a) Using methods of separation commonly employed to cut tape from the roll (e.g., hand-torn, scissor-cut),
b) Using readily available sources of tape (e.g., sold in retail stores for different end-uses),
c) Applying stretching of the fragments that simulate manipulation while binding a victim, or wrapping an object,
d) Examining tape pieces of sizes typically found in questioned items,
e) Using storage, preservation, and recovery methods often employed in crime laboratories to examine the evidence.

Assessing the performance of the method in the "validation set" provides an estimate of the performance in casework, including the sources of uncertainty and error rates. Some of the factors, however, could affect the performance of the method in unpredictable ways. To mitigate this, the validation set should consider different variables during the experimental design (dependent variables, independent variables, and controlled variables). The size of the dataset and the balance

of true negative and true positive samples are also factors that should be carefully selected during the study.

The data generated during the comparison of torn tape edges includes qualitative and quantitative information. For example, description of the overall edge pattern, color, texture, scrim orientation and construction, and alignment of backing and adhesive elements are qualitative data that is collected during the comparison of tapes. On the other hand, quantitative data is collected during the comparison of duct tapes in the form of a similarity metric (see section 2.2.4) and is documented in a scale of 0 to 100.

We have defined the edge similarity score (ESS) on duct tape fracture comparisons as the number of consistent scrim areas across a fracture pair by the total scrim areas along the width of the tapes, as shown in Equation 1 below:

$$\text{ESS} = \frac{\text{\# matching scrim bins}}{\text{total \# of scrim bins}} \times 100 \qquad \text{Equation 1}$$

While the ground truth of casework samples cannot be known, the origin of the samples in a validation set is often established to evaluate the performance of the method. As a result, the comparison of two tape ends can result in one of the following four outcomes:

1) True positive, which is a comparison pair that was once part of the same piece and is correctly reported by the examiner as a match.

2) True negative, which is a comparison pair that is known to originate from the separation of different areas of a tape roll that is correctly reported as a non-match.

3) False negative, which is a comparison pair that was once part of the same piece and is incorrectly reported by the examiner as a non-match.

4) False positive, which is a comparison pair known to originate from the separation of different areas of a tape roll that is incorrectly reported as a match.

Using the number of samples correctly reported or misclassified, performance rates such as sensitivity, specificity and accuracy can be calculated. The true positive rate for the method is known as the sensitivity, and is calculated as shown in Equation 2, where TP represents the number

of true positives of the results reported by the examiner, and FN denotes the number of false negatives:

$$\% \text{ Sensitivity } = \frac{TP}{TP+FN} \times 100 \qquad\qquad \text{Equation 2}$$

The true negative rate for the method is the specificity, calculated as shown in Equation 3, with TN representing the number of true negatives, and FP the number of false positives of the results reported by the examiner:

$$\% \text{ Specificity} = \frac{TN}{TN+FP} \times 100 \qquad\qquad \text{Equation 3}$$

False positive and false negative rates can be calculated using Equations 4 and 5, respectively:

$$\% \text{ False Positive Rate } = \frac{FP}{FP+TN} \times 100 \qquad\qquad \text{Equation 4}$$

$$\% \text{ False Negative Rate } = \frac{FN}{TP+FN} \times 100 \qquad\qquad \text{Equation 5}$$

The accuracy rate can be calculated using Equation 6, where IN represents the number of inconclusive results:

$$\% \text{ Accuracy } = \frac{(TP+TN)}{TP+TN+FP+FN+IN} \times 100 \qquad\qquad \text{Equation 6}$$

Accuracy rates for a method are beneficial to evaluate overall how likely the examiner was to make a misidentification. An important aspect when calculating the accuracy is to maintain a balanced dataset regarding true positive and true negative samples. The accuracy of the method will not be as reliable of a measure of the method if the dataset is unbalanced. For instance, with a larger number of true negatives than true positives, the overall effect of misidentifying a true positive will be masked by the number of true negatives identified.

Receiver Operating Characteristic (ROC) curves can be used to visualize the performance of the ESS classifier. ROC curves are plotted with the true positive rate on the y-axis, and the false positive rate plotted on the x-axis. The diagonal (y=x) is representative of a classifier that is

randomly guessing a class (positive or negative). A classifier under the diagonal is a classifier that performs worse than a random guess. A high performing classifier will follow the y-axis and then left-to-right the top axis of the graph. An example is shown in Figure 1.5, with the green line representing a high performing classifier, and the blue and red representing progressively poorer classifiers.
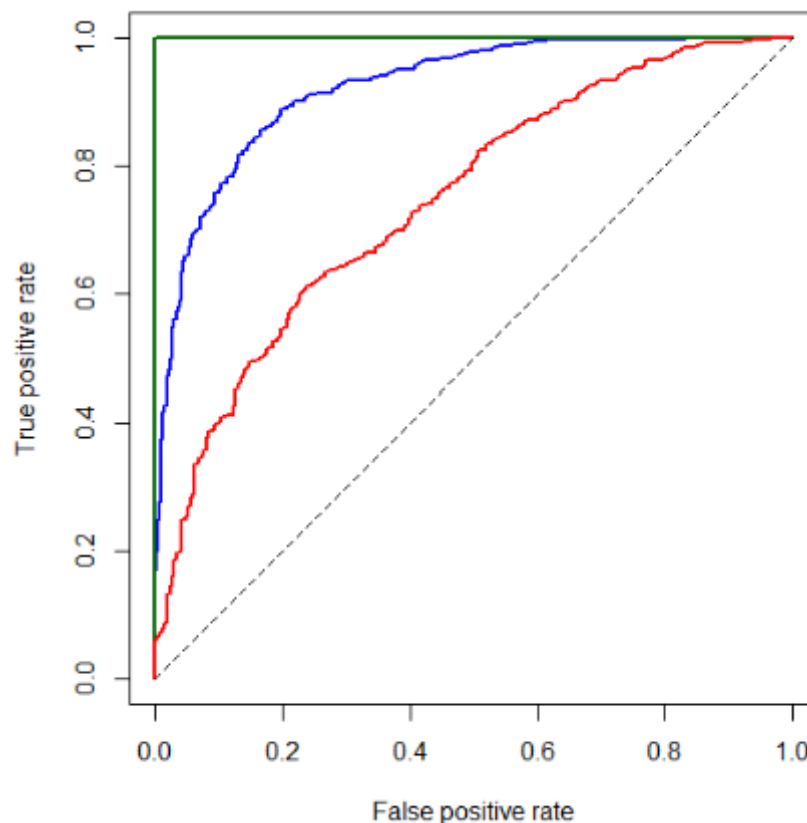
Figure 1.5. An example of ROC curves. The green line represents an ideal classifier, with the blue and red showing progressively poorer classification capability. Graph generated for illustration purposes with simulated data in R.

ROC curves are dependent on the separation between the positive and negative classes, but not the distribution of those classes. The area under the curve (AUC) is often used to reduce the ROC performance to a singular value. The AUC of a classifier represents the probability that the classifier will rank a randomly chosen positive with a higher score than a randomly chosen negative [48]. With a single classifier (ESS) being evaluated, the AUC can be used as an additional estimate of the overall accuracy of the method, as it is evaluating the accuracy at regardless of threshold for the dataset.

The similarity metrics and effect of different variables on the distribution between true positive and true negative samples can also be evaluated by exploratory tools such as boxplots, histograms and probability density functions. Boxplots are used to graphically describe numerical data through quartiles without making any assumptions to the distribution of the data. Boxplot can have extended whiskers that indicate data outside of the upper and lower quartiles. Quartiles divide a sorted list into four equal parts. The first quartile (Q1 or 25th quantile) is the middle number between the smallest number in the dataset and the median of the data set. The second quartile (Q2 or 50th quantile) is the median of the data. The third quartile (Q3 or 75th quantile) is the middle value between the median and the highest value of the dataset. The "box" of the boxplot is drawn connecting Q1, Q2 and Q3. The whiskers can either be defined by the absolute minimum and maximum values in the dataset, or using 1.5 times the interquartile range (IQR). The IQR is calculated by subtracting Q1 from Q3. Therefore, the lower whisker of the boxplot would be determined by the smallest value that is still within the range of Q1 – 1.5 IQR, while the upper whisker would be defined by the largest value that falls within Q3 + 1.5 IQR. Values that fall outside of the range including the IQR are considered outliers [49]. An example boxplot can be seen in Figure 1.6.
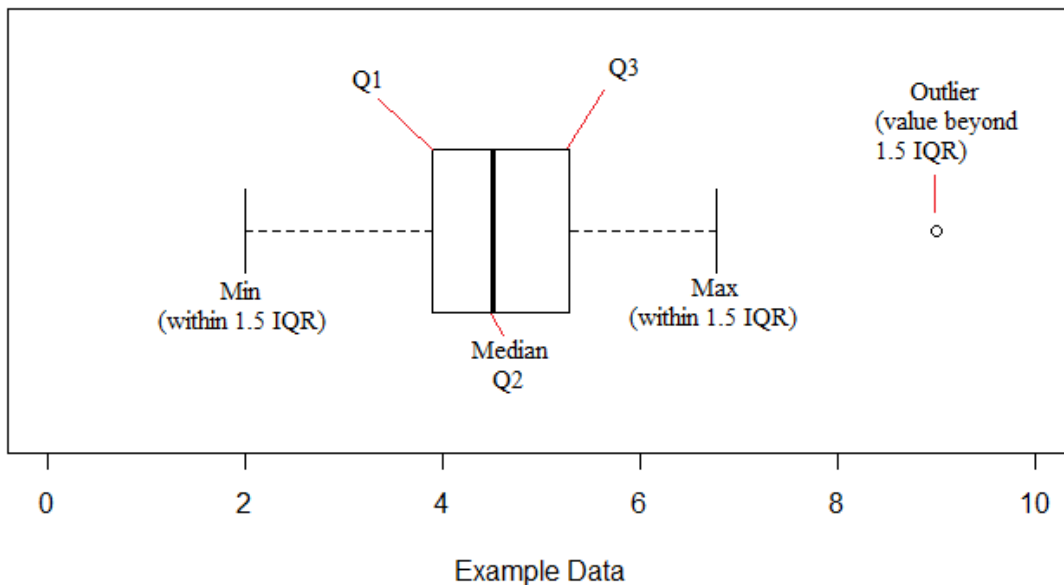
Figure 1.6. An example boxplot showing the various quartiles and values that are used to calculate the range of the box and whiskers. Graph generated for illustration purposes with simulated data in R.

Comparing datasets using boxplots can allow for the observations of many variables for the dataset. A wide range between Q1 and Q3, and therefore a wide range for the whiskers, indicates a high degree of variability within the dataset. If the range of whiskers or boxes of separate boxplots overlap each other, then the two datasets show similar distributions. However, if the datasets have a distinctive separation, it suggests that there are substantial differences between the datasets. An example is shown in Figure 1.7. The ranges of Sets 1 and 2 overlap, indicating there is not a substantial difference between the sets. However, neither set overlaps with Set 3, showing there is a difference between Set 3 and the other two. As boxplots are only exploratory tools, further significance testing should be done, however, to assess significant differences among groups.
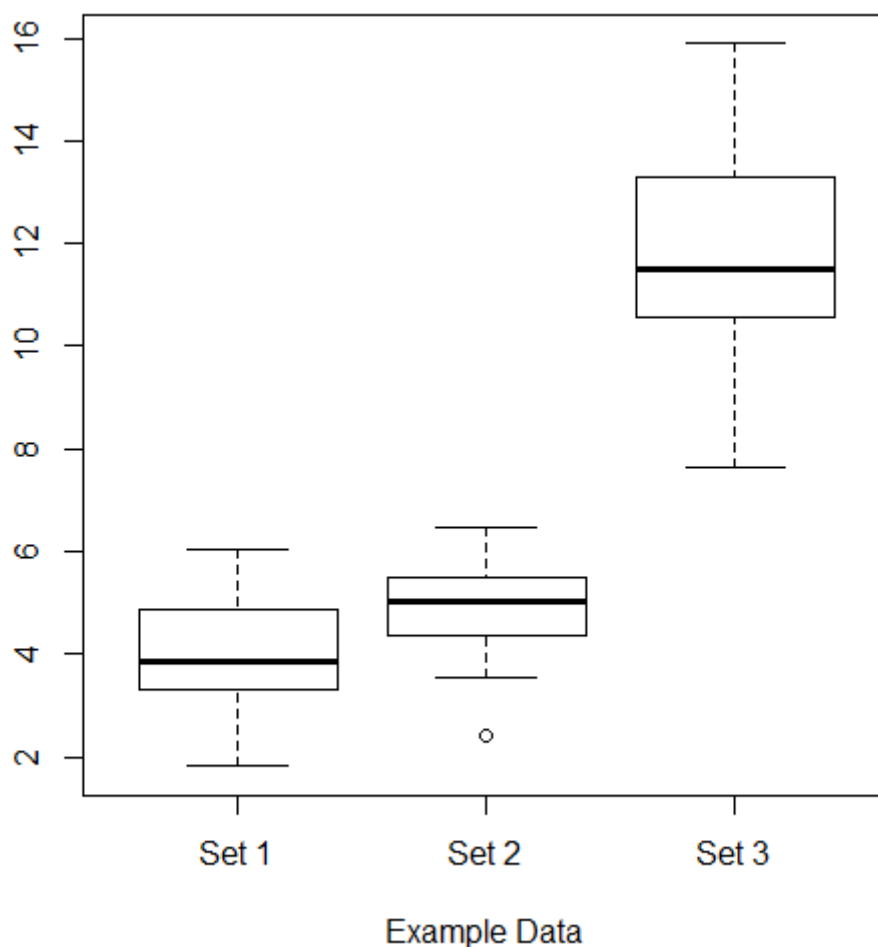
Figure 1.7. Three example datasets plotted as boxplots. Sets 1 and 2 overlap with each other but do not overlap with Set 3. Graph generated for illustration purposes with simulated data in R.

In addition to boxplots, frequency (or density) histograms can be generated from scores collected from true positives and negatives. Frequency histograms are generated by dividing the range of values into a series of intervals, known as bins. The quantity of values in each bin is the frequency for that bin. Bins are usually designed to be consecutive and non-overlapping to prevent data from being counted multiple times. The bin size is critical to the construction of histograms, as the distribution of the data can be misleading if the bin size is too large or too small. An example is shown in Figure 1.8, with the same 100 data points plotted with a bin width of 2, 5 and 10.
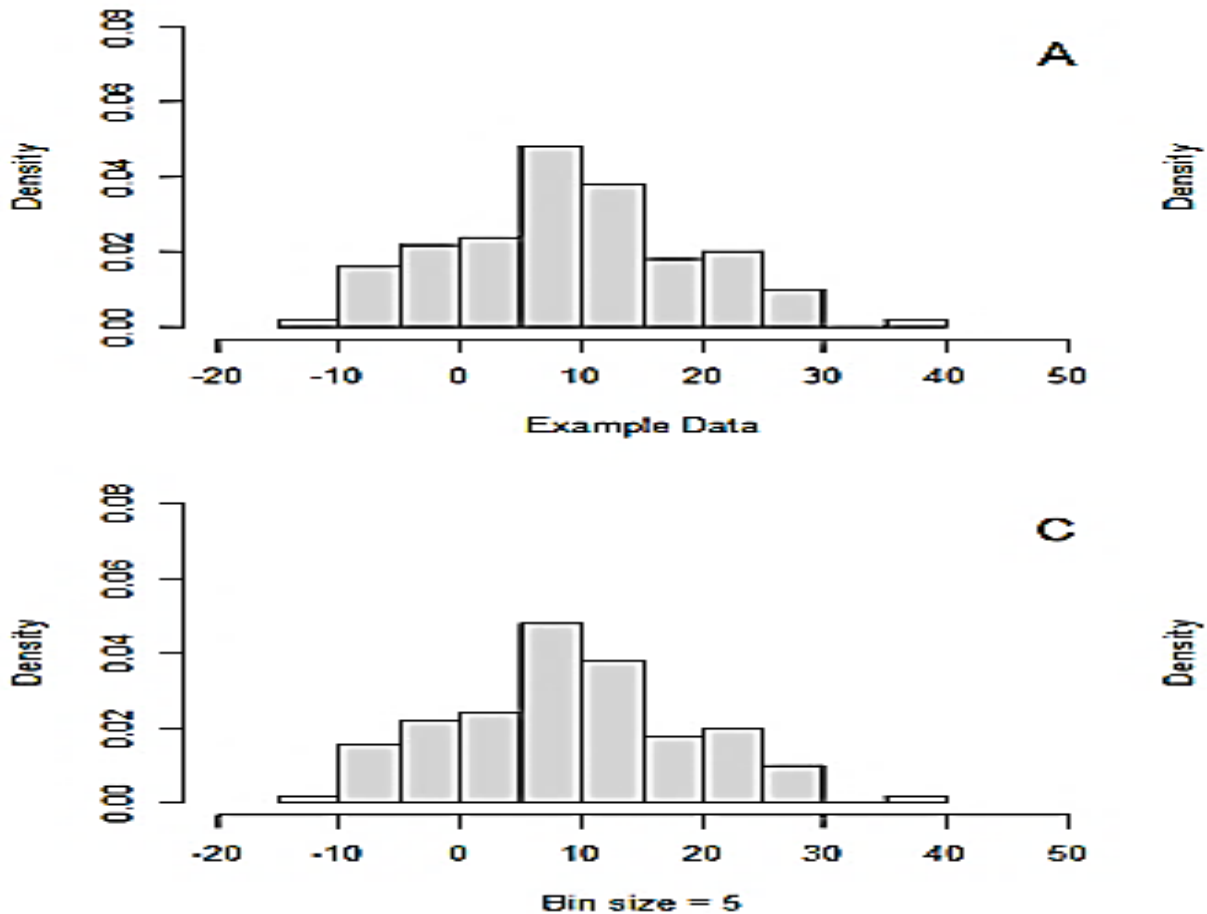
Figure 1.8. A series of histograms showing the same 100 random normally distributed data points. The original dataset is shown in image A, with images B, C, D demonstrating bin widths of 2, 5 and 10 respectively. Graph generated for illustration purposes with simulated data in R.

There is no specific bin size recommended – the most efficient bin width is dependent on the dataset [50]. It is possible to have multiple bin widths within the same histogram. Varying bin width is particularly useful when data is skewed or bimodal, as using wider bins where the

frequency of data is low reduces noise, yet using narrow bins where frequency of data is high offers greater precision to density estimation.

Using the histograms, a probability density function (PDF) can be estimated. A PDF is a function that describes the probability that a random variable will fall between a given range of known values. The x-axis for the PDF curve is the experimental data values, and the y-axis represents the density of those values. The probability is calculated by measuring the area under the PDF curve between the desired range of values. The integral of the entire PDF is equal to one. There are many ways to estimate PDF curves, but the most effective way is to estimate directly based off the data [51]. The histogram allows the PDF to be estimated, but also has the potential to lack continuity, as seen in image A, B and C of Figure 1.7, where there are no data points between 30-35. The lack of data may be corrected with the measurement of more samples, but that is not always a viable solution. While normal distributions can be used to estimate a general PDF, many times the data is not normally distributed, and so the estimated probabilities would not be reflective of the data. To this end, a non-parametric approach allows for the probabilities to be estimated directly from the data itself and allows the user to control the degree to which the curve fits the data. Just as the bin width is used to control the smoothing of histograms, the smoothing of a PDF curve can be controlled by a weighing function, or kernel. Some of the most common kernels include Gaussian, Triangular, Rectangular, Biweight, and Epanechnikov [51]. PDF estimations are controlled by the shape of the function (controlled by the kernel) and degree of smoothing for the kernel (controlled by the bandwidth). An example of Gaussian kernel density estimation is shown in Figure 1.9. The data is shown in histogram form in image A, with three different bandwidths applied. The bandwidth of 0.5 shows extensive under smoothing, with spiky curves closely following the shape of the histogram. Conversely, the bandwidth of 10 causes over smoothing, and some of the shape of the distribution tail around 40 is lost.

As an additional method to evaluate the score criteria method, a likelihood ratio (LR) is a statistical test used for the comparison of the goodness of fit of two models, one typically named as the null hypothesis (H_1) and the other one named the alternative hypothesis (H_2). In forensic science, likelihood ratios are used to represent the ratio of the probability of the evidence (E) given two alternative hypotheses and can be used as a means to quantify the strength of a conclusion [52]. A

LR is a ratio of the probabilities of observing certain evidence under two competing hypotheses, H_1 and H_2, often associated with the prosecution and defense, respectively.
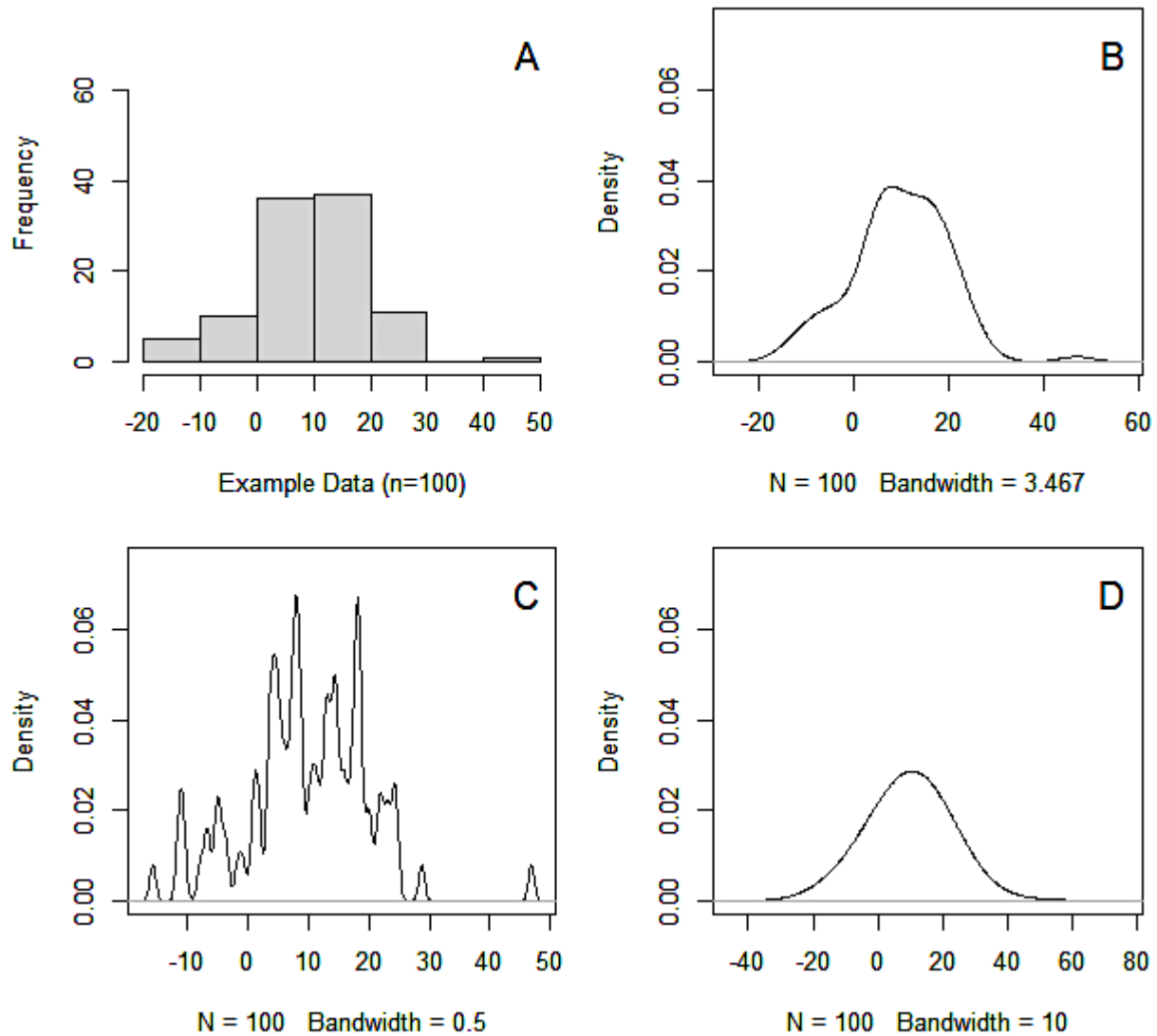
Figure 1.9. Kernel density estimations estimating the data in image A. Image B shows an effective bandwidth of 3.467, which shows the shape of the distribution without extensive loss. Images C and D demonstrate under smoothing and over smoothing of the data when the appropriate bandwidth is not used. Graph generated for illustration purposes with simulated data in R.

For example, in the case of a physical fit between duct tape edges, the likelihood ratio would represent the probability of the evidence (physical fit) given the hypothesis that the tapes originated from the same source, over the probability of the evidence given the hypothesis that the tapes originated from different sources. If the hypothesis is formulated based on source attribution, with

a likelihood ratio of one, the evidence does not provide enough information to support either hypothesis. A ratio above one supports the *same source hypothesis* while a ratio below one supports the alternative, *different source hypothesis*. The greater or smaller the number - on a relative scale - the stronger the support to a particular hypothesis [52].

Different forensic comparative analysis disciplines reported the use of a simplified approach to evaluate the weight of the evidence for handwriting, biometric data, fingerprints, and glass elemental profiles using score-based likelihood ratios (SLR) [53-56]. These methods use similarities or differences observed between samples. However, in this approach the "evidence" is limited to a "score" and modelled under two competing hypotheses.

Score-based models are generally reported using hypothesis such as:

H_1= "x" score resulted from the comparison of two fracture end patterns left after separation from the same source
H_2= "x" score resulted from the comparison of two fracture end patterns left after separation from different sources

Nonetheless, it is worth pointing out that LR and SLR cannot be considered as equivalent. Under LR framework, the question is: *How strongly do you believe the source of interest (Q_1) would have produced the edge fracture if it had been the true source?* With SLR the question formulated is: *How strongly do you believe the comparison of two fracture ends collected from the same sources would produce the same result (score) as when Q_1 is compared to the known source (K)?*

In a simplified equation, a SLR can be expressed as:

$$SLR = \frac{P[Score \mid Same\ Source]}{P[Score \mid Different\ source]}$$
Equation 8

Utilizing score-based likelihood ratios can reduce the overall features that are identified when reporting a score versus the whole assessment of the evidence characteristics [54]. The quality of a duct tape severed edge is highly dependent on a number of factors, including the quality of the tape, the method of separation, degree of damage to the tape edges during collection as well as

other factors. The degree to which these factors affect the quality of severed tape edges has been evaluated to some extent but is not fully characterized. The factors may also be linked, which further complicates the quality of the edge (i.e., a low-quality tape is more likely to be deformed by stretching). With those attributes, the calculation of the likelihood ratio is more difficult; for the likelihood ratio to be used, the factors that affect the tape edges and any links between them must be able to be characterized probabilistically [55]. However, using score-based likelihood ratios can overcome some of the difficulty. The examiner is able to estimate the degree of similarity or difference between tape edges with all factors considered but the distinctive features and quality of the match are reduced to a singular similarity score that is reported. Looking at an estimated distribution of SLR values can offer a practical insight into the weight of a particular similarity score reported during a physical fit comparison. However, because of the subjectivity by which the conclusion is made, there must be some caution in relying purely on a calculated SLR when determining the value of the evidence.

Histograms of LR (or SLR) values can be generated to assess the amount of misleading evidence within a dataset. Ideally, matching samples should receive high scores, corresponding to high SLR values indicating strong support for the conclusion of match. However, there is a realistic chance for overlap of SLR values. SLR values can be evaluated in a histogram as an alternative way to assess the discrimination power of the SLR values. Misleading evidence would provide a number of SLR under 0 for H_1, or SLR values higher than 0 under H_2 [57]. In general, the higher the range of overlap (ROA) the lower the accuracy of the method to differentiate between the groups. An example is shown in Figure 1.10, where with H_1 (same source) in blue and H_2 (different source) in green.

Rates of misleading evidence can also be demonstrated through Tippett plots. Tippett plots show the cumulative LR (or SLR) histograms. Two curves representing the probability of achieving an LR for H_1 and H_2 are plotted against the log10(LR) values [58]. The rates of misleading evidence are measured as the value of the curve at the value of 0 on the x-axis. Strong misleading evidence can cause the curves to shift and decrease the separation between the curves.
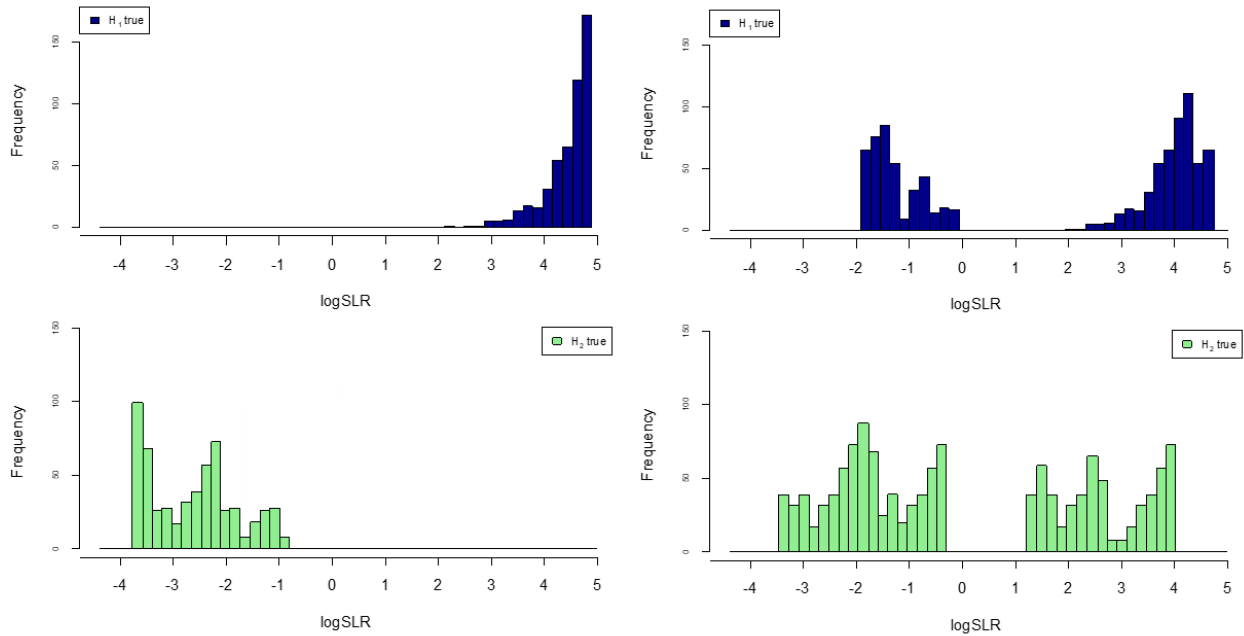
Figure 1.10. Figure plotting the log10 of the SLR for a dataset. SLR values under H_1 are shown in blue, and SLR values under H_2 are shown in green. Left graphs show ideal separation between SLR. Right graphs show example of misleading evidence as observed by overlap of the H_1 and H_2 distributions. Graph generated for illustration purposes with simulated data in R.

An example Tippett plot is shown in Figure 1.11. The misleading H_2 evidence is illustrated as area filled in blue and the misleading H_2 evidence by the area filled in red. An extrapolation to the y-axis of the point in which the straight line at LR is equal to one crosses the curve can be used to estimate the misleading rate.

A combination of the statistical methods described can be used to assess the value of a physical fit between duct tape edges. As such, examiners can choose the statistical method that provides the most information or use the ESS alone to interpret of the quality of the fit and support their conclusion.
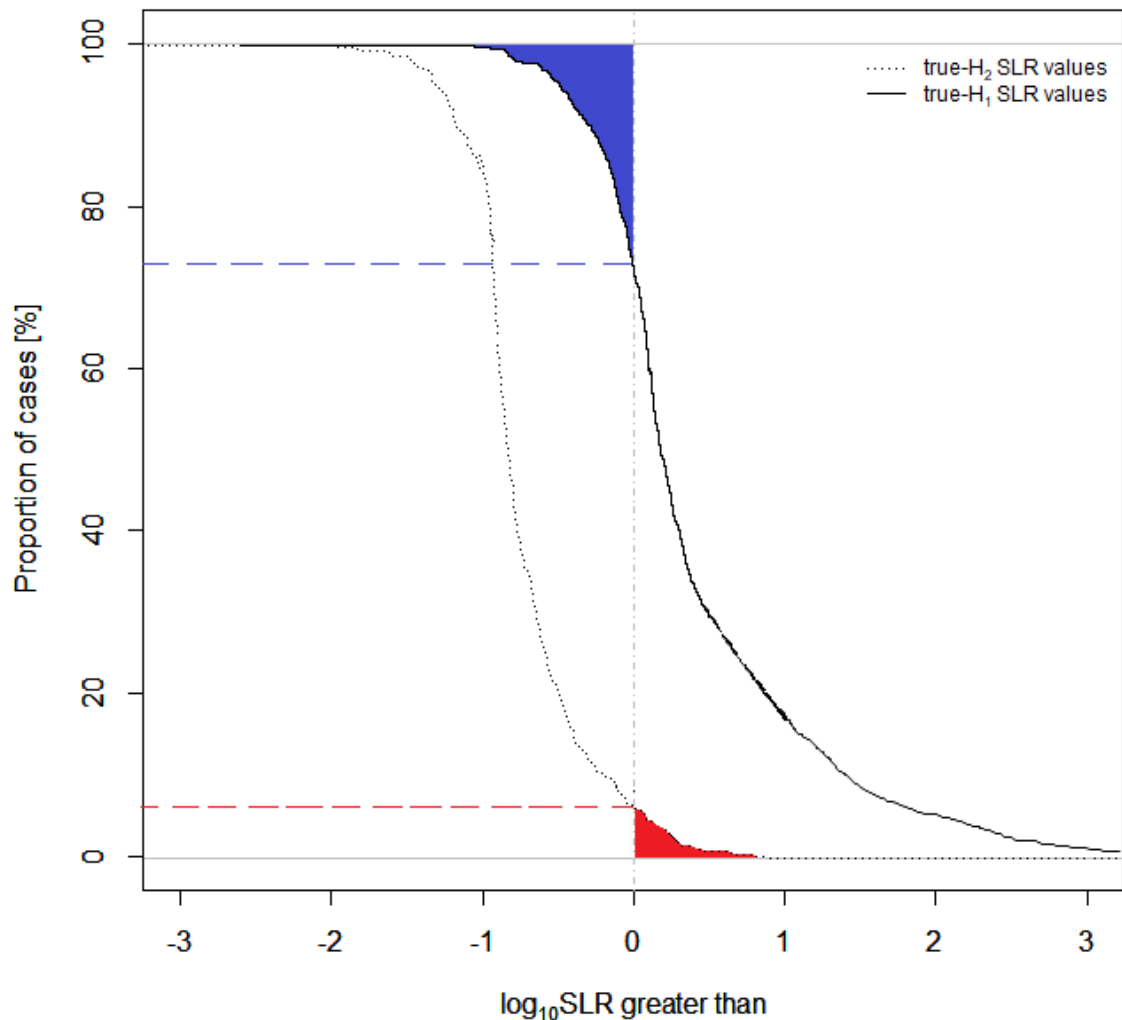
Figure 1.11. Example Tippett plot where the areas denoting misleading evidence marked in blue and red for H_1 and H_2 respectively. Graph generated for illustration purposes with simulated data in R.

1.5 Instrumental Analysis of Tape

The use of instrumental analysis in tape examination is fundamental to provide additional discrimination in cases where the physical features are consistent across samples but no fracture fit is identified. The chemical analysis permits classification of tapes based on the identification of components, even in the absence of a known source.

FTIR

One of the most common techniques is the use of Fourier Transformed Infrared Spectroscopy (FTIR). FTIR utilizes infrared radiation to interact with the sample. The absorption of the infrared

energy is measured as a function of changes in dipole moment and vibrational energy gaps – molecules can only absorb certain discrete frequencies. FTIR spectra provide information on the structure of the molecules and bonds within a sample. Reference spectra are used to identify the main components of unknown samples. FTIR spectrometers have a high signal-to-noise ratio and can quickly scan a wide range of infrared frequencies [56]. The technique is non-destructive, and therefore the samples of tape backing and adhesive can be used for further analysis. There are standards created by ASTM to guide the use of FTIR analysis for tapes, including general sample preparation, handling and the number of replicates that should be collected [59]. However, typical FTIR configurations used for infrared analysis of tapes offer bulk analysis. The analysis of the tape sample produces one spectrum for all the components of the sample, which can cause certain components to be masked.

Py-GC/MS

Another method used to analyze tapes is pyrolysis gas chromatography-mass spectrometry (Py-GC/MS). The instrument complements the information obtained by FTIR. During Py-GC/MS analysis, a small piece of tape (~50 µg) is heated very quickly in an inert environment. The organic constituents break down into smaller fragments, while the inorganic components are left behind. The pyrolized organic fragments are carried through a gas chromatograph, which separates the fragments according to their affinity with the column. The chromatographic step normally employs narrow columns to separate mixtures of organic compounds, and provide retention time for the components. In a regular mass spectrometer, the sample is ionized by electron impact, then transferred to the mass analyzer (quadrupole) where the analytes are filtered based on their mass-to-charge ratio, and the abundance measured [60]. While very informative, the Py-GC/MS is highly time consuming and destructive to the sample and therefore is conducted at the end of the analytical sequence.

LA-ICP-MS

Laser ablation inductively coupled plasma mass spectrometry (LA-ICP/MS) is an analytical technique that utilizes a laser beam to remove a small amount of mass from the surface of a sample of interest [61]. The interaction of the laser and the sample induces heating, evaporation and ionization of particles in the sample. The removed ions and particles are carried into an ICP-MS, where the nanoparticles are ionized in the core of an argon plasma, and the ions are analyzed by

their mass-to-charge ratio. This method offers high sensitivity and selectivity, with limits of detection in the low and sub-µg/g range [12,14]. The technique is minimally destructive, as only a few micrograms of material are removed at a time. However, LA-ICP-MS instruments are, in general, costlier to acquire and therefore less available in forensic laboratories than SEM-EDS and XRF systems.

SEM-EDS

Scanning Electron Microscopy Energy Dispersive Spectroscopy (SEM-EDS) can be used to provide information about the elemental composition of backings and adhesives. It is currently the standard method for the elemental analysis of tape samples [2]. The technique measures the energy or wavelength characteristic X-rays of the elements present within the sample. SEM utilizes an electron beam to interact with the sample, pushing electrons from the atomic orbitals by inelastic collision. In response, electrons from higher energy orbitals will transition to lower energy levels and in the process emit an X-ray. The energy of the emitted X-ray is characteristic of the particular element. Characteristic X-rays are generated from electronic transitions of electrons in the orbital shell. As the electrons move from a higher energy state (such as a L or M shell electron) to a lower ground state, the difference in energy emitted in the form of a photon. An example is shown in Figure 1.12. Atoms can only undergo certain transitions, which allows the discrete energy of the escaping photon to be used to identify the atom.
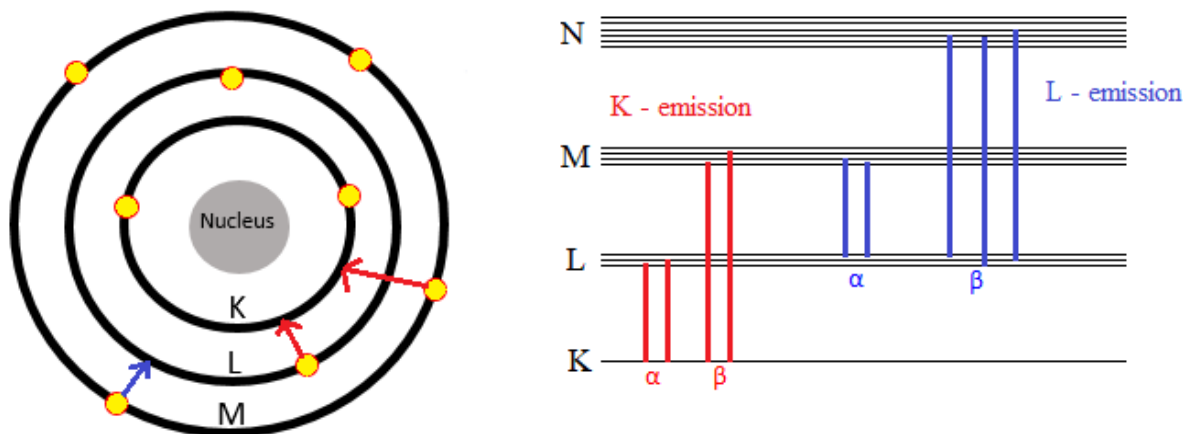
Figure 1.12 Potential transitions for shell electrons. The difference in the energy of the starting state and the ending state is the energy of the emitted photon.

The electron beam can be modified by apertures, lenses and electromagnetic coils to control the intensity and diameter of the beam [62]. The emitted X-rays are measured by the detector as the number of photons detected at each discrete energy value. The photon counts are translated into a continuous spectrum with the counts on the y-axis and the energy or wavelength values on the x-axis. The most common type of detector is energy dispersive spectroscopy (EDS), which simultaneously measures the energies of all photons coming from the sample.

There is also wavelength dispersive spectroscopy (WDS), in which the number of photons is measured at specific wavelengths. Wavelength dispersive spectroscopy is more selective and more sensitive than energy dispersive spectroscopy, but the instruments utilizing WDS are generally more expensive, so many labs utilize EDS detectors exclusively [63].

SEM-EDS also offers the advantage of imaging at high magnification. While this feature is not as commonly needed for tape samples due to the larger size of most trace evidence tape samples, having access to the images can help prevent measurement of contaminants on the surface of the tape. SEM-EDS is a surface analysis technique, as the electrons only penetrate a few nanometers into the sample. As a result, there is very little sample preparation required for tape samples. Compared to LA-ICP-MS, SEM-EDS is less sensitive and selective, but is more commonly found in forensic laboratories.

XRF

X-ray fluorescence is a spectrochemical technique used for the elemental analysis of materials. The technique is very similar to SEM-EDS, with the exception that XRF uses photons (with energy up to 50kV) instead of an electron beam to generate the characteristics X-rays. The most common source for XRF instruments is an X-ray tube. Within the tube, current is run through a tungsten filament to generate electrons. The electrons collide with the anode material (usually Mo or Rh) and eject photons that are directed at the sample. It is important when conducting XRF analysis to know what the source is made with, as that element cannot be measured in the analysis [64]. Background radiation known as Bremsstrahlung radiation is generated due to interactions of electrons with each other and the atomic nuclei [64]. Most software used for analysis of XRF data can automatically reduce Bremsstrahlung radiation. Other X-ray sources exist, such as

radioisotopes, but their use is dwindling in favor of the cheaper and safer tube source. Due to the higher energy, XRF penetrates deeper (μm to mm) into the sample. The penetration depth depends on the energy of the element, and the density of the sample [60]. The equation to calculate the depth of penetration is seen in Equation 9, where I is the quantity of photons returning from the sample, I_0 is the quantity of photons entering the sample, μ/ρ represents the mass attenuation coefficient of a given element for a particular matrix, and x represents the density of the object [64]:

$$\frac{I}{I_0} = e^{\left(-\left(\frac{\mu}{\rho}\right)x\right)}$$
Equation 9

XRF instruments are likely to penetrate through tape samples, due to their relatively low density. Therefore, more sample preparation is required to separate the backing and adhesive layers, and then mount them on surfaces that do not produce interferences. XRF instruments can also be modified with components such as polycapillary lenses to control the energy and intensity of the beam and filters act to maximize sensitivity to elements of interest by reducing some of the background scatter and the intensity of lower energy peaks.

Two of the most common detectors are silicon lithium (or SiLi) and silicon drift detectors (SDD). With SiLi detectors, the characteristic X-rays interact with the surface of the detector, and release electrons that travel towards the back of the detector. The drop in voltage is proportional to the energy of the X-ray. SiLi detectors must be cooled regularly, because the detector can break if the temperature gets too hot. SDD detectors do not require liquid nitrogen cooling, as the cooling is accomplished by the Peltier effect, in which the electrons are slowed as they move through rings of alternating material. Other kinds of detectors, such as Scintillation and Gas detectors are used in XRF instruments but are more commonly used in wavelength-dispersive spectroscopy. [63,64].

The parameters of an XRF instrument can be optimized for the detection of major elements in a sample. In general, increasing the beam sampling area and the extending the acquisition time can be used to improve the signal to noise ratio. The voltage applied is also critical. For a characteristic peak to be detected, the sample must be struck with a beam that is at least 1.5 times the amount of energy of the x-ray transition. The working distance must also be controlled, which is the distance between the end of the x-ray source and the sample. If the working distance is off, the sample may

not be in focus. As a result, there would be a decreased ability of the characteristic x-rays reaching the detector, and an increase in background noise. The main components are depicted in Figure 1.13.
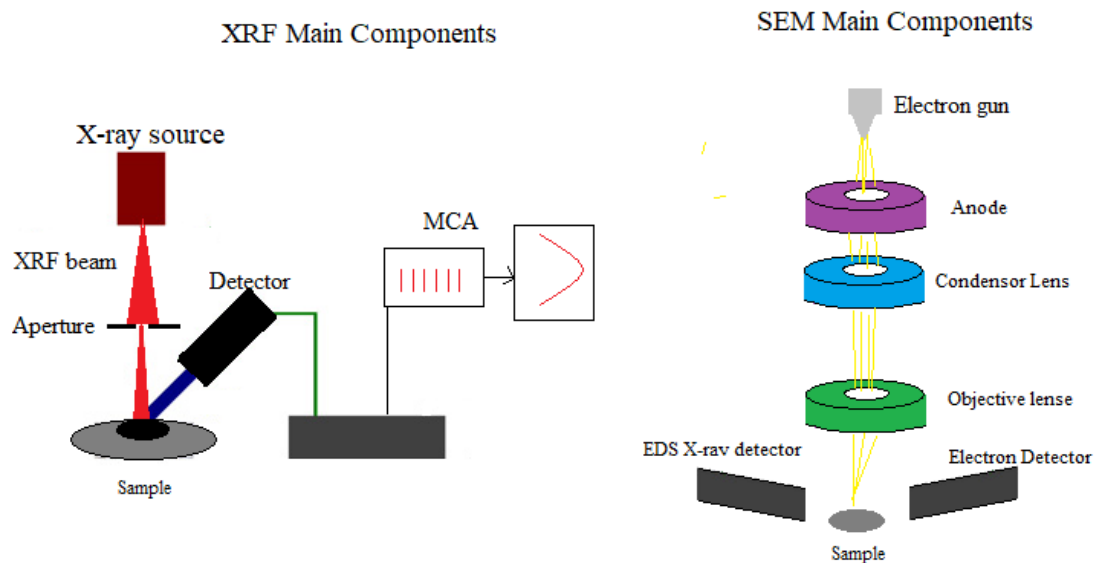
Figure 1.13 Main components of XRF and SEM instruments.

A series of recent studies have focused on the examination of black electrical tape specimens with a variety of analytical techniques, including FTIR, SEM-EDS, and Py-GC/MS and LA-ICP-MS. The major components present in electrical tape were determined, and the discrimination power of each method was evaluated for each instrument or a combination of techniques. Of the methods evaluated, the LA-ICP-MS had the greatest discrimination at 94% of the electrical tape backings, with SEM-EDS as the second most discriminating at 87% [12,14,61]. Although SEM-EDS is widely available in crime laboratories, LA-ICP-MS is more complex and expensive and therefore less accessible to the average forensic laboratory. As a result, this research aims to evaluate the utility and performance of XRF in comparison to SEM-EDS and LA-ICP-MS, as a more sensitive, yet affordable, technique.

1.6 Evaluation of Spectrochemical Data

Spectral overlay is commonly used in forensic spectrochemical analysis to quickly distinguish samples of differing composition. Spectral overlay is the process of plotting the spectra of a questioned sample on top of the spectra from the known comparison sample in order to determine differences — or lack of — in composition between the samples. Analyzing multiple replicate

measurements from different areas of the questioned sample and from different pieces of the known sample allow the analyst to assess the within-sample variability.

During the overlay of the spectra, the examiner evaluates any difference between two samples including presence or absence of characteristic peaks, or differences in the relative heights and shapes of the peaks. If multiple measurements within each sample are collected, the analyst can establish a min-max range of the variation of each peak of interest. If the ranges of the known and questioned sample do not overlap, then the samples are distinguished. Otherwise, if none of the peaks in the spectra differ, the samples are considered indistinguishable (Figure 1.14).
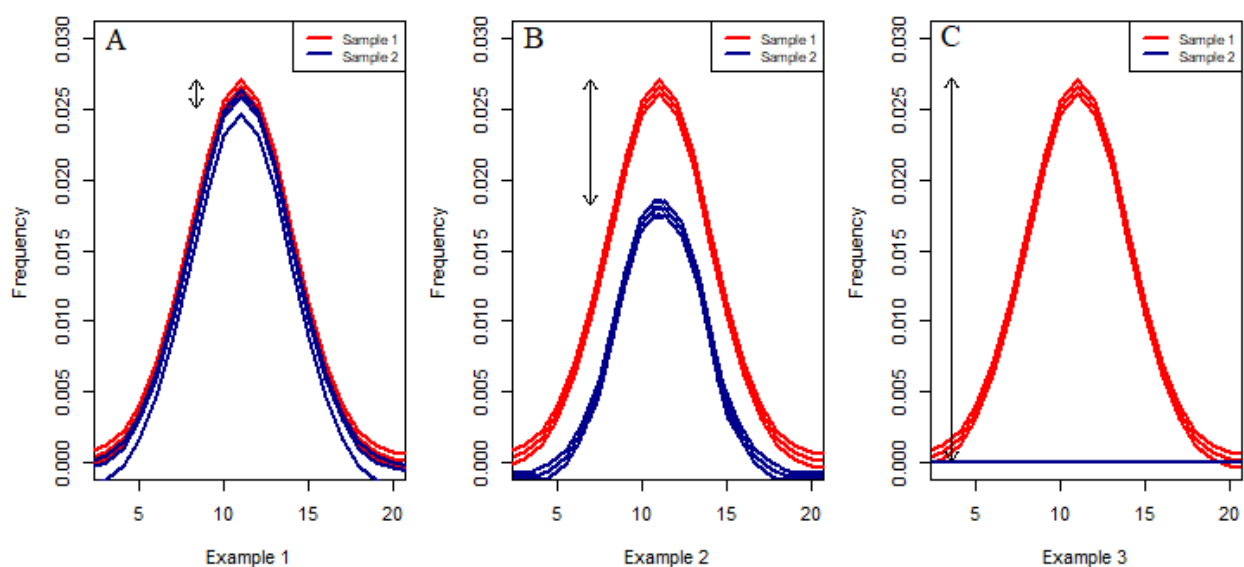
Figure 1.14. Examples of spectral overlay. Image A shows two sample spectra overlapping – making them indistinguishable at this peak. Image B shows both peaks above detection limit, but with a significant difference in intensity that would allow them to be distinguished. Image C shows the absence of a peak for one sample, but the presence for the other sample, this would also allow the samples to be differentiated.

Using spectral overlay, examiners can pre-screen the spectra and observe distinctive differences in the elemental profile of the samples. However, spectral overlay is a more subjective approach, as the analyst still must determine whether or not there are significant differences in peak height and shape – a concern particularly with XRF data due to the higher proportion of background noise present in the spectra and the potential for a lack of resolution between element peaks of similar energy. As such, it was important to develop a semi-quantitative statistical approach to support decisions based on spectral overlay observations and aid in differentiating between samples,

particularly those that were otherwise indistinguishable by spectral overlay. The developed method was based on the study by Ernst et. al. for forensic glass samples that was incorporated into ASTM E2926-17 [66,67]. The integrated areas under the peaks and signal-to-noise ratios were calculated and compared statistically utilizing Analysis of Variance (ANOVA) and Tukey-Kramer post-hoc tests. For peaks to be used in the semi-quantitative analysis, they must be above a SNR of ten to meet the requirements for compliance – also known as the limit of quantitation (LOQ) [66].

ANOVA is a statistical approach used to compare means between independent groups. ANOVA tests show if at least two samples in the group are significantly different from each other, and is particularly useful with balanced data. ANOVA can be used to summarize regression with many predictors [68]. This is particularly helpful with electrical tapes, as there is no guaranteed composition for electrical tapes and so a predictable series of peaks cannot be used in the data analysis. The Tukey-Kramer test (also known as Tukey-Honestly Significant Difference) follows the ANOVA test, and performs each pairwise comparison of sample means, and calculates a HSD statistic to identify specifically which pairs are significantly different from each other. Tukey-Kramer is a more conservative post-hoc test, ensuring that if there was a difference in samples that it would not be a coincidence that the ranges were significantly different.

2. Objectives

The main goal of this study was to assess the reliability of physical end matching and chemical comparison of pressure sensitive tapes, and to provide forensic examiners more objective criteria to inform and support their conclusions.

This research aimed to expand and enhance the current methods by which pressure sensitive tapes are analyzed and interpreted in forensic laboratories. In that regard, the methodologies applied in these studies are in alignment with current protocols and instrumentation available in crime laboratories to facilitate their future incorporation.

The research centered on addressing current needs for two of the tapes most frequently received as trace evidence: duct tapes and electrical tapes. The manufacture and composition of duct tapes and electrical tapes impart very different physical and chemical properties in these two end products, requiring different attention. For instance, duct tapes are composed of a reinforced fabric between the backing and the adhesive layers. The scrim fibers impart resistance during the tearing, keeping a several features in the fractured edges that may be valuable during a physical match analysis. Also, the physical properties such as color, thickness, width, scrim count, and construction have shown to be the most discriminating features during the examination.

On the other hand, electrical tapes are primarily composed of an adhesive and a flexible backing that easily deforms during the separation from the roll. Unlike duct tape, the physical properties of the backing and the adhesive are not very informing. However, the chemical composition, and in particular the elemental composition of backing layers, provides good discrimination among different sources.

As a result, the first part of this research was focused on establishing the basis to demonstrate the scientific validity of physical fit examinations on duct tapes, while the second part of the study was directed to improving the chemical examination of electrical tapes. (Figure 2.1) Specifically, the primary objectives of this research were to:

1) Objective 1: Develop and validate methods to evaluate the reliability of physical end matching of **duct tapes**.

2) Objective 2: Evaluate the performance of elemental analysis of **electrical tape backings** by XRF and spectral comparison methods.

Assessing the reliability of physical end matching and chemical comparison of pressure sensitive tapes

Objective 1: Evaluate the reliability of physical end matching of **duct tapes**	Objective 2: Evaluate the performance of elemental analysis of **electrical tape** backings by XRF and spectral comparison methods
Tasks Task 1. Develop a standard protocol for the quantitative assessment of the quality of a physical fit between duct tapes. Task 2. Assess the standard criteria on various tape samples and calculate threshold values and performance rates. Task 3. Assess the evidential weight of a physical fit through statistical analysis on similarity scores.	Tasks Task 1. Develop and validate an XRF method for the analysis of electrical tape backings. Task 2. Compare XRF performance to SEM-EDS and LA-ICP-MS and evaluate data from multiple XRF configurations Task 3. Assessment of spectral data with various spectral comparison methods.

Figure 2.1. A summary of the overall objectives and tasks of the research project.

The following tasks were completed throughout the project to accomplish the research objectives (Figure 2.1):

Task 1 (Objective 1) – Develop a standard protocol for the quantitative assessment of the quality of a physical fit between duct tapes.

To date, there have been no scientifically-estimated criteria or standardized protocol for examiners to utilize when finding if two edges realign with enough features to arrive at a match conclusion. In that respect, the first task of this research involved defining what features were relevant and valuable for the comparison of duct tape edges, what the smallest sub-unit of comparison should be, how qualitative features should be documented, and how to translate qualitative observations

to a quantitative value. For this purpose, the scrim areas or "bins" were selected as the smallest comparison unit. The comparison bin, therefore, was defined as the torn area between two of the warp (lengthwise) scrim fibers. The hypothesis was that since scrim areas in duct tape are a consistent feature across a tape roll, they provide a systematic mean to ensure the examiners observe the same areas and the same number of regions when performing an examination. This systematic comparison of the scrim areas allows for more straightforward documentation of the quality of the fit, and therefore a demonstrable training and peer review process. The tapes were observed under the stereomicroscope at 10-40x, allowing documentation of relevant features such as backing dimples and markings, adhesive and backing protrusions and indentations, unsymmetrical scrim fibers across the tear, and tear edge feature directionality. These features contributed to determining if the scrim bins were found to "match". The edge similarity score (ESS) was reported as the number of scrim bins that aligned across the edge as a relative percentage of the total number of scrim bins along the total width of the tear. Using the edge similarity score allows the criteria to be applied to a variety of tapes, regardless of the absolute number of scrim bins within each tape.

Task 2 (Objective 1) – Assess the standard criteria on various tape samples and calculate threshold values and performance rates.

With the comparison protocol developed, a total of 2280 comparison pairs were blindly evaluated over the course of the study to determine what characteristics of the tapes would have the most significant effect on the performance of the method and the distribution of the similarity scores. Three tape grades - low, mid, and high-quality - were evaluated. Approximately 200 comparison pairs were used to evaluate the error rates for each of the low and high-quality tape sets. Conversely, the mid-quality tape was utilized to assess the effect of the separation method and stretching on the error rates. A total of 1804 comparison pairs were generated using the mid-quality tape. Separation method was evaluated through 500 pairs generated from scissor-cut samples and an additional 508 pairs generated from hand-torn samples. To assess the effect of stretching on the performance, moderate stretching was applied to hand-torn samples to create another set of 500 pairs. Additionally, tape used to bind a victim in a mock homicide case was presented to the examiners to create a scenario of extensively stretched samples (288 comparison pairs total). The complete breakdown of the sets used are seen in Figure 2.2.
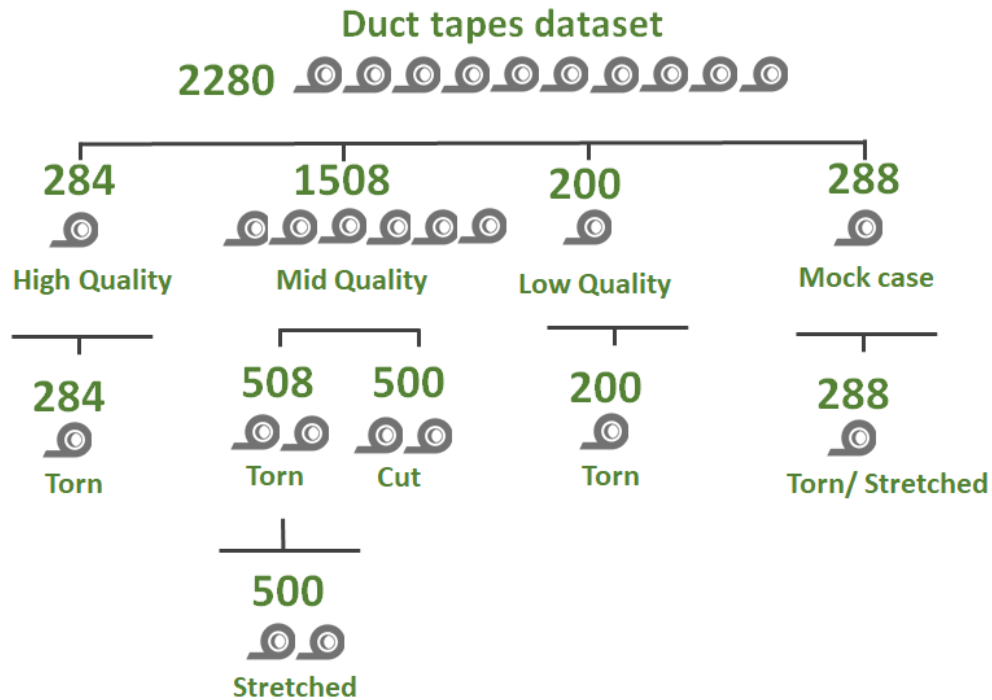
Figure 2.2. A breakdown of the total number of comparison pairs evaluated throughout the course of the duct tape physical fit study.

The examinations were performed by five student examiners who underwent training to perform the comparisons, with pre-defined criteria to master the training as further explained in chapter 3. Sample identification numbers were randomly generated, and the examiners were kept blind to the origin of the samples until all the comparisons were completed, and the scores reported to avoid bias. From the similarity score results, performance rates including sensitivity, specificity, false positive and false negative rates, and the overall accuracy were calculated. Threshold values were estimated by assessing the distribution of scores of true positive and true negative tapes.

Task 3 (Objective 1) – Assess the evidential weight of a physical fit through statistical analysis on similarity scores.

While the similarity score can be used alone to describe the quality of a fit between edges, it is necessary to understand how any given score relates to a distribution of scores among a population where the known truth is known. The frequency distributions of scores were used to estimate the empirical probability of observing a true positive or true negative result for any given score. The

estimated probabilities can be used in a number of different ways to assess the criteria and the discrimination power of the method. The probabilities were used to calculate score-based likelihood ratios. The distribution of ESS and the SLR values can be used to estimate what characteristics of the fractured edge or the tape may have the most significant effect on the quality of the fit, and to assess the weight of the value of the evidence. Tippet plots were also calculated to evaluate the discrimination power of the method further. The combination of statistical methods offers additional support for examiners performing tape examinations. The addition of SLR values and Tippet plots allow the numeric value of a particular score to be translated into a degree of certainty for the conclusion of match or non-match.

Task 1 (Objective 2) – Develop and validate an XRF method for the analysis of electrical tape backings

While XRF has been used to determine the most discriminating elemental components of glass, the most relevant components of electrical tape have yet to be identified. There are no reference standards available for analysis of tape. While some elements can be predicted based on the most common components in electrical tape, there is no guarantee the elements will be present in every tape sample, nor present in a predictable concentration. To identify the relevant elements, spectral data from 89 black electrical tape backings were collected on the XRF. Three replicates were analyzed on each sample, and additional duplicates were monitored as quality control. The peaks for each spectrum were identified to demonstrate which elements were the most discriminating between different samples. To this end, the major discriminating elements were found to be Al, Si, Cl, Ca, Sb, Ba, Ti, Fe, Zn, Pb, Br, and Mo. Other elements such as Cd and Cr were also found in a few samples, but were not as common. The identified element peaks were used in the subsequent data analysis to identify their analytical performance, the instrumental variation, intra-roll, and inter-roll variation.

Task 2 (Objective 2) – Compare XRF performance to SEM-EDS and LA-ICP-MS and evaluate data from multiple XRF configurations.

The current standard method for the elemental analysis of tape backings is SEM-EDS. However, SEM-EDS is not as sensitive of a technique as XRF, and can be more time-consuming, and costlier. LA-ICP-MS offers higher sensitivity and selectivity and lower limits of detection when compared to SEM-EDS and XRF. However, LA-ICP-MS instruments are not as common in forensic

laboratories compared to SEM-EDS and XRF systems. Many forensic laboratories already utilize XRF as an analytical technique for its ease of operation, non-destructive nature, and rapid analysis. XRF instruments are also relatively less costly to acquire than SEM-EDS or LA-ICP/MS instruments. As such, the performance of XRF was compared relative to SEM-EDS and LA-ICP/MS for the elemental analysis of 40 electrical tape backings. The resulting groups of distinguishable samples were assessed to compare the discrimination power for the methods. In addition to comparing the different techniques, several configurations of XRF instruments were used to assess the performance of the method, including selectivity, reproducibility, and discrimination capabilities. The three configurations included an IXRF beam coupled to an SEM-EDS (referred to as XRF A), as well as two standalone XRF instruments, a Thermo ARL QUANT'X EDXRF (XRF B) and a Bruker Tornado XRF (XRF C). Figure 2.3 shows the breakdown of the number of samples and the total number of comparison pairs within each set.
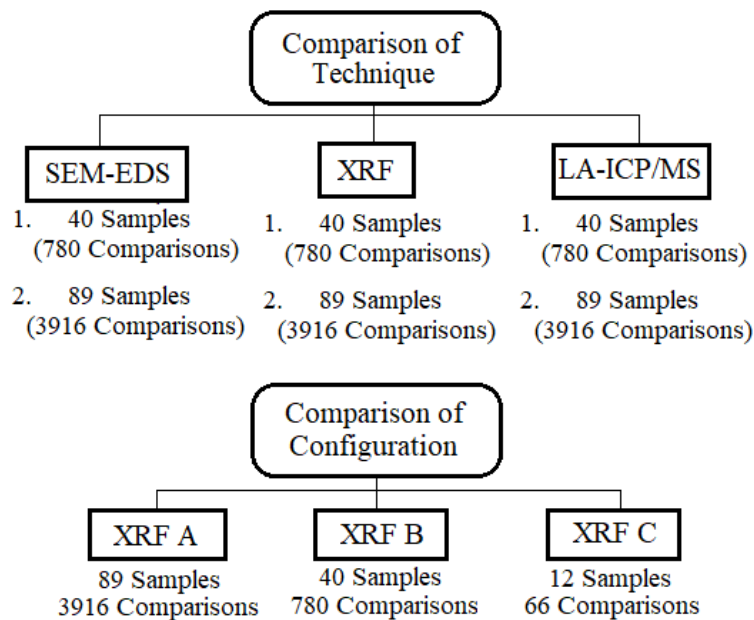
Figure 2.3. Diagram representing all the sample breakdown and the total number of comparison pairs within the three techniques and the three XRF configurations utilized in the course of the study.

Task 3 (Objective 2) – Assessment of spectral data with various spectral comparison methods.

Two approaches were taken with regard to the collected spectral data. The first consisted of a qualitative spectral overlay between comparison samples. Spectral overlay stands as the current method for comparing spectral data from XRF instruments, as samples of different composition can be rapidly distinguished and commercial software permit the overlay of multiple samples at

once. If at least one element in the spectra did not overlap in the peak range for the compared samples, the samples are considered as non-match (originating from different sources). This can take the form of the presence or absence of a characteristic peak in one sample, or differences in the relative heights and shapes of the peak between replicates of two samples. As such, spectral overlay is a valuable screening technique, particularly when samples have distinct differences in composition. However, there is a degree of subjectivity to spectral overlay, particularly when two samples have very similar compositions and peak heights and shapes. In addition, spectral overlay can be highly time-consuming when there is a large number of samples. A more objective, semi-quantitative method based on the ASTM E2926-17 method for XRF analysis of glass samples was established to support the observations made using spectral overlay [65,66]. The method includes the integration of areas under the peaks, the calculation of signal-to-noise ratios, and utilization of Analysis of Variance (ANOVA) and Tukey-Kramer post-hoc test to statistically evaluate differences in composition between samples.

Detailed results and experimental design for Objective 1 are provided in chapter 3, while Objective 2 is expanded in chapter 4.

3. Development and validation of a systematic approach for the quantitative assessment of the quality of duct tape physical fits

3.1 Summary

Duct tape is a common type of evidence found at crime scenes such as sexual assaults, murders, kidnappings, and bombings. During the examination of a known and questioned item, a 3D realignment along their edges is known as a physical fit and is often regarded as conclusive evidence that the items were once part of a single object. The conclusion of a fit between edges relies on the examiner's judgment to identify distinctive features across the tape ends. However, there are currently no consensus-based methodologies or standards to inform their opinions. This study developed a practical method to qualify and quantify tape end match features using edge similarity scores (ESS) and provided an empirically demonstrable basis to assess the significance of duct tape fracture fits. ESS were calculated as a relative ratio of observed matching sections per scrim bins across the fractured edge, providing a quantifiable criterion and means for a systematic peer review process. A set of 2280 duct tape end comparisons were analyzed for the validation study. The weight of physical fits was evaluated through similarity metrics, probabilistic estimates from the distribution of ESS, and score-based likelihood ratios. The effects of separation method, stretching, and tape grade on the distribution of ESS and the overall accuracy are reported. The low and mid-quality tape sets had low misidentification rates, with 0% false positives and 1-2% false negatives. The accuracy of the low-quality hand torn set was 99.5%. For the mid-quality tape, the scissor-cut set had an accuracy of 99.8%, whereas the pristine and stretched hand-torn sets both resulted in an accuracy of 99.6%. The high quality hand-torn set had a more significant number of misidentifications, with 21.4% false negatives, 0% false positives, and an accuracy of 84.9%. On average, ESS higher than 80% provided a score-based likelihood ratio (SLR) that supported the conclusion of a match, and ESS lower than 25% provided an SLR supporting the conclusion of non-match.

3.2. Introduction

Duct tape is a common type of evidence received in forensic laboratories as it is used in a variety of criminal activities. There is a high discrimination potential for duct tapes based on the variability

of physical characteristics, such as backing color and texture, adhesive color, tape width and thickness, and fabric construction [4]. Duct tape, unlike other kinds of tape, contains a layer of fabric enforcement known as the scrim layer. The scrim makes duct tapes more resistant to deformation and stretching than other tapes with no reinforcing structure such as vinyl electrical tapes or packaging tapes [41]. As such, duct tape samples may exhibit a higher potential for fracture fits.

A physical fit examination is performed during analysis of trace evidence, where the examiner compares the edges or faces of a questioned item to a potential known source. Examiners attempt to determine a source by identifying distinct tear patterns and features often left during the separation of the materials. Fracture fits represent the highest degree of association between questioned materials [36,37]. In the absence of a physical fit, analysts conduct a full chemical analysis of the tape to determine if the pieces could be distinguished by other class characteristics or composition [3]. Failure to identify differences in the known and questioned samples would conclude an association with class characteristics between the samples. While some laboratories require full characterization of the tapes regardless of the results from a fracture examination, most stop further analysis if a fit is identified. As such, false positives represent a higher risk for misleading results as, unlike a negative or inconclusive physical fit conclusion, a "match" result will not undergo additional examination steps.

Previous studies have evaluated the ability of examiners to correctly identify fracture fits for materials such as glass, plastic, metal, paper, paint, and tape [25-28, 33, 35, 36, 41, 68]. However, only a few have addressed performance measures, such as error rates [35-37]. In studies where the ground truth is known, there are five potential outcomes for physical fit comparison. True positive outcomes occur when samples that originated from the same source are reported correctly, whereas samples correctly identified as originating from different sources are labeled as true negatives. False positives are samples incorrectly identified as originating from the same source, and false negatives are samples incorrectly identified as coming from different sources. An inconclusive result indicates there are not enough distinct features to conclude match or non-match. This information can then be used to estimate performance rates, such as sensitivity, specificity, and accuracy.

A few studies have assessed examiner accuracy for duct tape physical fits and evaluated what factors of the tape (e.g., grade, quality, thickness, separation method) could affect accuracy. A study by Bradley et. al. had four examiners comparing a total of 81 hand-torn and scissor-cut tape ends. Another examiner blindly reviewed edges that were not correctly identified. The authors found that 92% of the tapes were identified correctly, with no false positives reported [36]. McCabe et. al. conducted a comprehensive study of duct tape examination, with three student examiners comparing a total of 2400 tape pairs, comprised of eight different duct tapes separated by hand, scissors, box knife, or Elmendorf tear tester. The participants reported a score defined by the relative length of the fracture with matching features respective to the entire width of the sample [35]. However, the estimation of the length of a fracture is somewhat arbitrary because a tape fracture is rarely straight. Overall, the study found accuracy for the torn tape to be 98.58-100%, and for cut tapes to be 98.15-99.83%. They found false positive rates to be under 3.33%, and false negative rates to be under 2.67%. Despite collecting the scores for each tape, no other statistical analysis was reported to assess the weight of the evidence [35]. Both studies observed that the quality and separation method had some influence on the conclusions. Namely, it was found that lower quality tapes and cut tapes were more likely to have higher percentages of false positive and false negative results [35,36]. However, no statistical assessment was reported to support the observation. The score as defined by McCabe et. al. was tested at the beginning of this study, however, it was quickly determined that there were many discrepancies between how the examiners evaluated the length of the edge, particularly in tapes that did not have a straight edge or had multiple, non-continuous areas of misalignment. As such, a modified methodology was needed to mitigate discrepancies between the examiners.

Regardless of the probative value given to a physical fit, the examiner does not have scientifically-estimated criteria to decide if the edges are similar enough to be considered a match. Studies conducted on duct tape samples to date have focused on qualitative methods of comparison. In the absence of standard thresholds and known error rates, the identification of tape fracture fits will remain subjective, and its scientific validity will likely be challenged in court. Therefore, there is a critical need to develop standardized match/non-match criteria, and to assess the validity and accuracy of tape fracture fit determinations using those criteria. This study is anticipated to strengthen current practice by providing a systematic method for the quantitative comparison of duct tape edges. The method supports the examiner's opinion by a more transparent interpretation

of the data, a more straightforward peer review process, and quantitative and empirical basis that substantiates the evidential value of fracture fit conclusions.

3.3. Materials and Methods

Three brands of duct tape purchased at local hardware stores were used: Duck Brand Electrician's Grade Gray Duct Tape (Duck Brand, ShurTech Brands, Avon, OH), Tool Bench Hardware Duct Tape (Greenbrier International, INC., Chesapeake, VA), and Silver Gorilla Tape (Berry Global, Evansville, IN). There are no standard criteria for determining the grade of tape, so for this study, the cost, package grade labeling, and physical characteristics of the tape rolls (scrim counts, the thickness of the adhesive and backing) were used to identify the tape quality. Table 3.1 lists the sets used in this study. The sets are referenced by their ID for the remainder of this manuscript.

3.1.1. Training

Before conducting experimental trials, the examiners underwent training in duct tape end matching. The student examiners received training discussing the history, manufacturing, and composition of duct tapes, and reviewed literature discussing physical matching and duct tape physical comparison. Each student then completed a training set consisting of at least 40 hand-torn and scissor-cut tape comparisons. The students reported their ESS for each comparison, as well as qualitative conclusions using one of five categories [Match+, Match-, Non-match+, Non-match-, Inconclusive]. The + represented a high degree of certainty, and the – indicated a low level of certainty of the conclusion. If incorrect categories were assigned, a remediation plan was developed that included the discussion of results with the instructor, identification of the sources of a discrepancy, and taking an additional training set until the examiner proved proficiency. A difference in ESS lower than 10% versus "ground truth" values reported by an expert was used as the criteria to declare the student passed the practical training. No significant discrepancies were experienced between the five students involved in the training.

3.2.2. Sample preparation
The samples were prepared by non-participating students, keeping the examiners blind to the origin of the samples during the examination and analysis of data. The samples were cut or torn

Table 3.1. Description of the tape sets and the number of comparisons in each set. In the identification of the tapes the following label convention was used: MQ = mid-quality, HQ = high-quality, LQ = low-quality, SC = scissor-cut, HT = hand-torn, and S = stretched.

Tape ID	Quality	Separation Method	Product	Backing Thickness (mils)	Adhesive Thickness (mils)	Scrim Count (warp/weft)	Number of comparison pairs	Number of examiners
MQ-SC	Medium (pristine)	Scissor-Cut	Duck Brand	4.0	2.5	20/8	500	1
MQ-HT	Medium (pristine)	Hand-torn	Duck Brand	4.0	2.5	20/8	508	2
MQ-HT-S	Medium (stretched)	Hand-torn	Duck Brand	4.0	2.5	20/8	500	1
LQ-HT	Low (pristine)	Hand-torn	Tool Bench	5.0	0.7	17/8	200	2
HQ-HT	High (pristine)	Hand-torn	Gorilla Tape	9.7	6.3	58/16	284	1
Case Study	Medium (stretched)	Hand-torn	Duck Brand	4.0	2.5	20/8	288	2

from the roll in pieces approximately four inches long and individually placed onto transparent acetate film (ACCO Brands, Booneville, MS). A random number generator was used to create sample ID numbers to avoid bias. The participating examiners were given a blank template to fill out with the comparison pairs pre-listed (Microsoft Excel spreadsheet). The samples in sets MQ-SC, MQ-HT, and MQ-HT-S were assessed using both edges on the tape, resulting in a total of four comparisons per tape pair. As a result, the sets had an unbalanced number of true matches and true non-matches. Nonetheless, this experimental set-up represented typical casework where the examiner evaluates the two edges of each known and questioned pair, and a maximum of one potential match is feasible among the four-pair comparisons. Sets LQ-HT and HQ-HT were labeled with only one comparison edge per pair, and overall had similar numbers of true matches and non-matches pairs.

The low and mid-quality tape sets had semi-transparent adhesive, allowing the scrim to be seen through the acetate and the adhesive. In contrast, the high-quality tape set had a thick, grey adhesive that did not allow for the scrim to be seen. As a result, during the comparison of high-quality tape, the adhesive was removed in a thin strip from each edge. Before removing the adhesive, the examiner observed the edges of the tape pair to note any features that would be beneficial in the final comparison in the event those features shifted or vanished. Then, the examiner lifted the tape edge from the acetate using liquid nitrogen and used a small quantity of hexane and cotton swabs to pull the adhesive away from the backing carefully. This method was used so to avoid shifting the scrim as much as possible.

3.2.3. Comparison Method

For each comparison pair, the examiners observed the physical characteristics, including the color of the adhesive and backing, the tape width, and the scrim count and pattern. The same roll of tape was used for the entire sub-sets, and so the physical characteristics were consistent throughout. However, the examiners were kept blind as to whether the roll was the same.

After completing the physical observations, the examiners put the joint edges together and observed the tear patterns under a stereomicroscope (Leica EZ4, PA, USA). For each tape comparison, the examiners were required to report a similarity score. The edge similarity score is defined as the number of scrim areas that align from one edge to the other as decided by the examiner, respective to the total number of scrim areas across the tape edge. Using the scrim bins as the smallest comparison units assures the examiners look at the same areas and the same number of regions when performing an examination, providing a systematic comparison protocol. The scrim areas were selected because the number of scrim fibers is a persistent feature of the roll. As shown by LaPorte and Wiemer, the scrim count remains constant, with variations of only ± 1 throughout the entire roll [5]. As such, the number of areas in between the scrim fibers represents a consistent subunit when viewing the edge of the tape. While measuring the relative length of the tape was initially considered as in [35], the measurements can be variable if there is any stretching or if the fracture edge is not entirely straight, and therefore the metric becomes somehow arbitrary. In contrast, estimating a score by scrim bins applies to a large variety of tapes, regardless of their number of scrims, which often changes with tape grade.

An example edge is shown in Figure 3.1, where 15 scrim areas are visible. The examiner would count the number of scrim areas across the entire width of the tape; this is the denominator. Then the examiner counts the number of matching areas and divides by the total scrim areas to calculate the percent of matching areas or ESS. The score is then utilized to make quantitative assessments of the weight of the evidence and provide a way to use a consistent reporting criterion across examiners. In this example, the ESS is estimated as 11 matching areas out of 15 scrim bins, or 73.3% ($\frac{11}{15} \times 100$). This is an example of a weak match (M-), where the stretching in areas 10 to 13 is a potential factor of the low ESS.

In terms of documentation, the examiner took notes of any significant features holding weight in their decision. In our experience, incorporating detailed documentation of the comparison edge features by scrim bins has proven effective during the peer-review protocol, added transparency to the decision-making process, and facilitated standardization of procedures. After the comparisons were completed and reported, the sets were compared to the master list with the

"ground truth" information to determine any misidentifications, as well as to evaluate the distribution of scores and the overall performance rates.
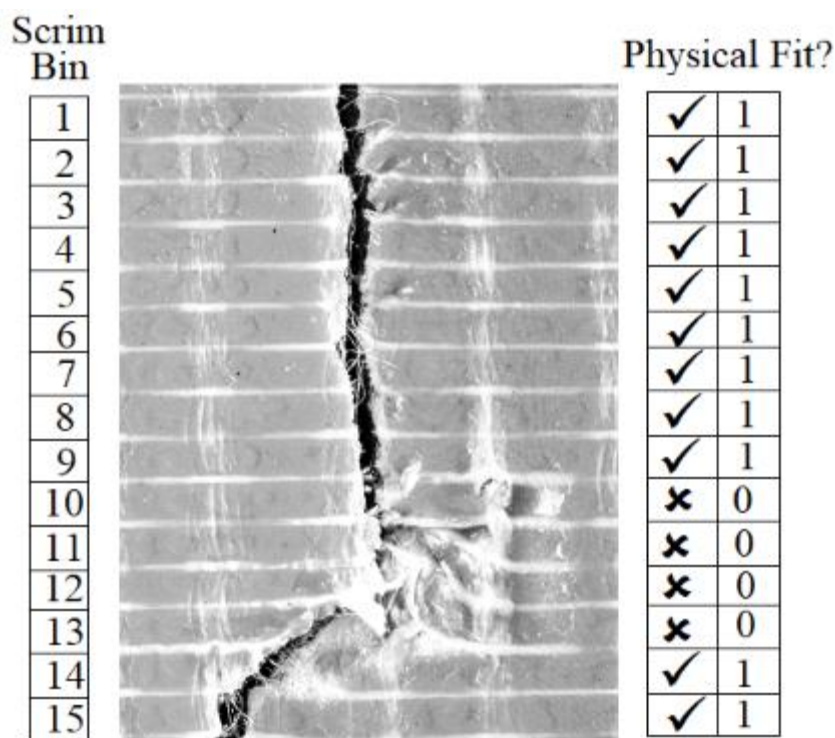
Figure 3.1. A section of duct tape showing the areas between the scrim that are used to calculate the ESS match score. An example template of what the examiner would record for each area is shown on the right, with check marks and the number 1 denoting matching scrim bins, and x's marking and the number 0 denoting non-matching scrim bins.

3.2.4. Performance Rates and Statistical Analysis

Performance rates were estimated for the sets such as sensitivity, selectivity, and accuracy as described in Section 1.4. In addition to the performance rates calculated for each examiner, the frequency distributions of scores were used to calculate score-based likelihood ratios (SLR). Various forensic materials, including handwriting, biometric data, fingerprint, elemental glass profiles, and footwear impressions have reported the use of a simplified approach to evaluate the weight of the evidence using score-based likelihood ratios [52,53,55]. These methods use similarities or differences observed between samples. However, in this approach, the "evidence" is limited to a "score" and modeled under two competing hypotheses. Hypothesis one for a duct

tape fracture fit comparison represents the probability of observing a given ESS if the two edges separated from the same source, while hypothesis two would be the probability of observing a given score if the two edges separated from different sources. The equation to calculate SLR is given in section 1.4. An SLR value equal to one would indicate that there is not enough evidence to suggest either a same source or different source conclusion. As the SLR values increase above one, the stronger the support for the same source conclusion, whereas an SLR less than one supports the conclusion that the objects originated from different sources.

The score frequency distributions were produced in Microsoft Excel (version Excel 2016). The scores for all tapes were recorded in Excel, and then the frequency function was used to classify the number of scores that fell within each bin that were true positives, true negatives, false negatives, false positives or inconclusive (bin size of 5%). The kernel density functions, probability estimates, ROC curves, SLR values, and Tippet plots were calculated using R software (RStudio version 1.2.1335, R version 3.6.0). To evaluate the score frequency distributions, and calculate the SLR values for each set, the scores for true positive and true negative tapes were imported into R and separated by their respective ground truth label. The R density() function was used to calculate the kernel density function for the TP and TN distributions. To estimate the probability, the density() function was integrated at each score value under the relative true positive or true negative distribution of the set. The probabilities were then used to calculate the SLR values for all comparison edges in the set, including the scores of any misclassified tapes. The frequency distribution of the natural logarithmic SLR values was used to generate the Tippet plots. The SLR frequency histogram plots and Tippet plots were generated based on R code derived from Zadora et. al. [56]. This process was repeated for all sets analyzed in the course of the study.

The frequency of occurrence on our data spanned from scores of "0" to "100". Some score bins, however, did not contain any empirical data within a set. Therefore, a probability of 0.0001 instead of zero was used in cases in which no empirical data was observed for a given score. Otherwise, SLR would be unrealistically reported as either zero or infinity. This probability was derived on the idea that comparing a more significant number of samples would eventually lead to a given data-score describing a matching tape and a non-matching tape.

3.2.5. Mock Case Study

A mock case study was designed to allow the examiners to evaluate tapes coming from a crime scene. The mock homicide case included a victim who was bound and gagged with duct tape, as seen in Figure 3.2A. The samples were collected in a manner that would cause more difficulty for the tape examiners – a few were placed on full acetate sheets, while some were crumpled up in plastic bags, as seen in Figure 3.2B. The tapes placed on the mannequin were torn from the roll. Known samples were cut from the edges of the rolls and placed onto acetate sheets.

Figure 3.2. Tape samples collected from the mock homicide case. Image A shows how the tape was used on the victim, while image B shows one of the collected tape samples that was crumpled in a plastic bag.

The questioned samples were given to the examiners to process and analyze for the edge comparisons. The origin of the samples was kept blind during the examinations. There were several tools available for the examiners to use in separating the edges of the tape, including a freezer, applying hexane, using a heat gun or liquid nitrogen, or a combination of methods. After all the edges were removed and placed onto clear acetate, the examiners were able to begin the comparisons to the known edges using the same method as for the reference sets.

3.3. Results and Discussion

The ESS provided a mechanism to qualify and quantify tape end match features and to estimate from empirical data the error rates and uncertainty associated with fracture fits. A total of 2280 duct tape edge-pair comparison were analyzed to calculate the occurrence of false positives and

false negatives (Table 3.1). Overall, the method allowed for high accuracy rates, with a false positive rate of 0% for all the tape sets evaluated.

The study evaluated the effects of separation method (hand-torn (HT), and scissor-cut (SC)), tape grade (low-quality (LQ), mid-quality (MQ), and high quality (HQ)), and stretching level (pristine vs. stretched samples (S)) on the overall accuracy and distribution of edge similarity scores.

Table 3.1. Performance rates for different tape sets analyzed.

Performance rate (%)	MQ-SC	MQ-HT (Analyst A)	MQ-HT (Analyst B)	MQ-HT-S	HQ-HT	LQ-HT
False positive rate (FP)	0.0	0.0	0.0	0.0	0.0	0.0
False negative rate (FN)	1.0	1.0	2.0	1.0	21.4	0.0
True negative rate (Specificity)	100.0	100.0	100.0	100.0	100.0	99.0
True positive rate (Sensitivity)	99.0	98.0	98.0	99.0	69.4	100.0
Accuracy	99.8	99.6	99.6	99.8	84.9	99.5

The following sections describe the characteristics of each set and compare their results and distribution of scores.

3.3.1. Medium Quality Tape Set

3.3.1.1. Scissor Cut Tape Set – MQ-SC

Set MQ-SC was made by cutting 125 pieces of tape. 100 of those tapes were randomly selected to make a total of 500 comparison pairs, of which 401 were true non-matches, and 99 were true matches. The tapes were cut in a single cut, either straight or at an angle, using scissors. Examples of true matches from the set are shown in **Error! Reference source not found.**.

Performance rates for the set are listed in **Error! Reference source not found.**. Only one pair was misidentified as a negative when it was actually a match.

Example 1.
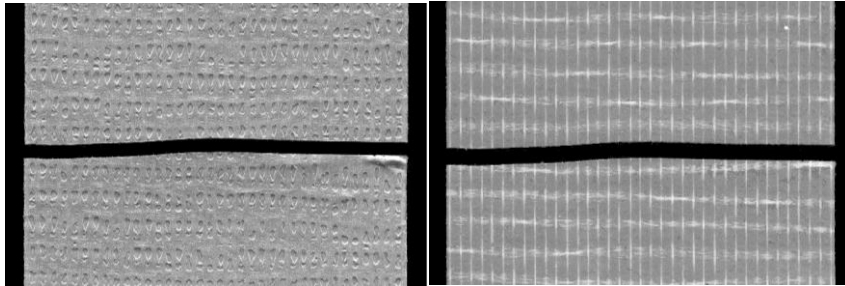
Example 2.
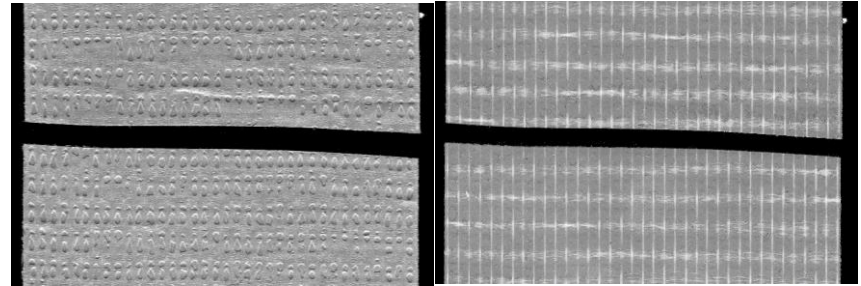
Example 3.

Example 4.

Example 5.

Example 6.

Figure 3.3. Six examples of true match comparison pairs from set MQ-SC. Left of each pair: backing side, right of each pair: scrim adhesive side.

Table 3.2. Relative performance rates for the MQ-SC set.

MQ-SC	Reported Non-Match	Reported Match	Reported Inconclusive	Total
True Non-Match	401 (out of 401, 100%) **True Negatives**	0 (out of 43, 401%) **False Positives**	0 (out of 401, 0%)	401
True Match	1 (out of 99, 1%) **False Negatives**	98 (out of 99, 99%) **True Positives**	0 (out of 99, 0%)	99

The accuracy rate for this set was at 99.8%. A concern for this set was that the separation of the tape by scissors would leave minimal distinctive features. However, despite the straight fractures, the remaining features were still sufficient for the examiners to correctly identify which tapes were true matches with a high degree of certainty. The true positives in this set were all given ESS of 85% or higher, as seen in Figure 3.4. In contrast, the true negatives received scores 40% or lower, with more than half the negatives receiving scores between 0-10%.
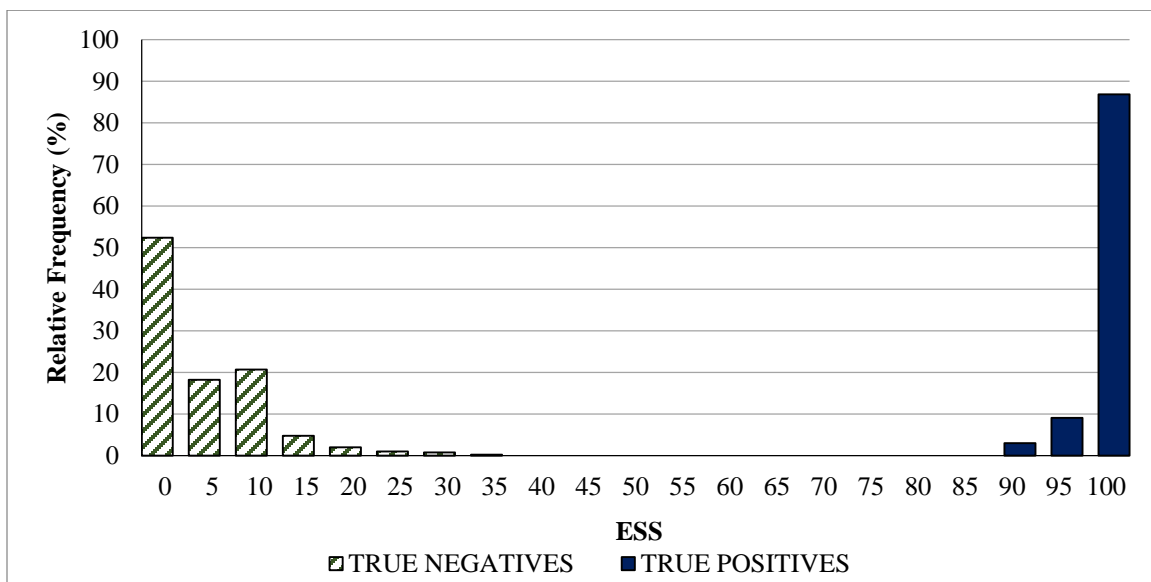
Figure 3.4. Relative frequencies of the ESS for the 500 comparisons completed for the scissor-cut tape set (MQ-SC).

Overall, there was a high rate of accuracy for the MQ-SC set. The relatively straight features left by the scissors did not impede the examiner's ability to make correct determinations as to the origin of the tape samples. This was in part due to the ability of the examiner to see through the adhesive to the scrim. In some cases, tapes that otherwise had straight fractures with very minimal features could be distinguished by orientation of the scrim fibers themselves – if the scrim did not

align across the tear, then very few of the areas would be in alignment for the score calculation. This feature required the examiner to be vigilant in positioning the tapes properly at the top and bottom, and ensure the position was maintained while counting the scrim areas.

3.3.1.2. Hand Torn Tape Set – MQ-HT

Similar to set MQ-SC, set MQ-HT was made using 100 tapes randomly selected out of a total of 125 tapes torn in a single motion. In total out of 508 pairs, 409 comparisons were true non-matches, and 99 were true matches. The set was observed independently by two separate examiners to compare the level of agreement on condition and ESS. Examples of true matching pairs are shown in **Error! Reference source not found.**.

Upon completion of the MQ-HT set, three main patterns were identified in the appearance of the torn edge. The qualitative proposed descriptions of the potential fracture patterns are described in Figure 3.6, labeled as angled, wavy and puzzle-like. The angled pattern consisted of a relatively straight tear from one end of the tape to another, without crossing over the tape backing. A wavy tear occurred as a curvy pattern found on more than 50% of the tape width, and puzzle-like edges included distinct cut-outs and protrusions.

After the examiners completed the comparisons, they were asked to review the edges of each tape and report the pattern. The most common pattern among torn tapes was angled (42%), followed by wavy (35%) and puzzle-like (22%). These descriptions were used to investigate how the edge pattern correlated with the incidence of false positives and false negatives.

The performance rates for the set are reported in **Error! Reference source not found.**, with both examiner's results. Two false negatives were reported in this set. The examiners misclassified the same two tape pairs. One pair was reported as an inconclusive by examiner A while examiner B reported the same pair as non-match. The other tape pair was reported as a non-match by both examiners. It was determined that the two pairs had small amounts of stretching on one edge, leading to incorrect conclusions. Both pairs were also wavy patterns.
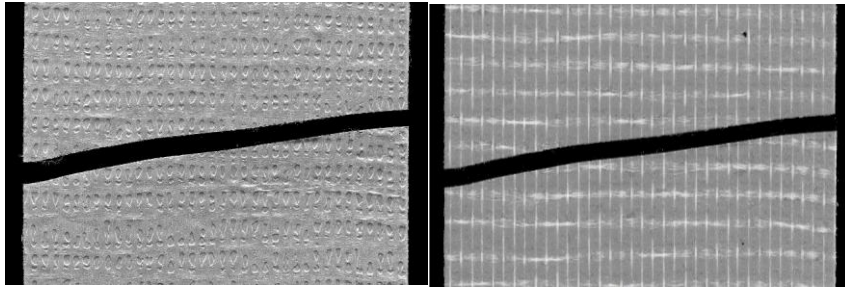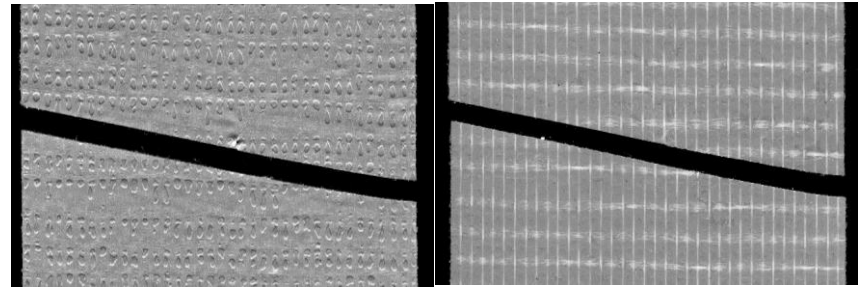
Example 1.
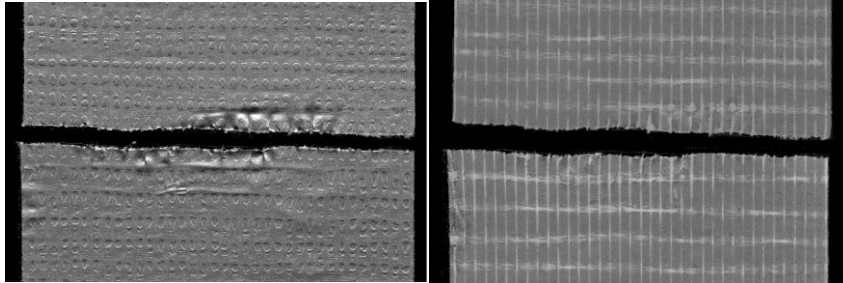
Example 2.

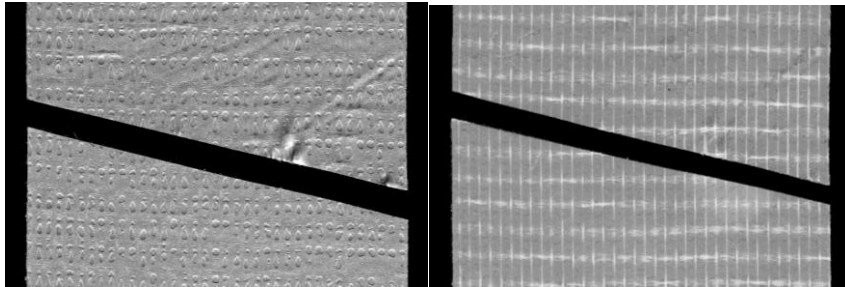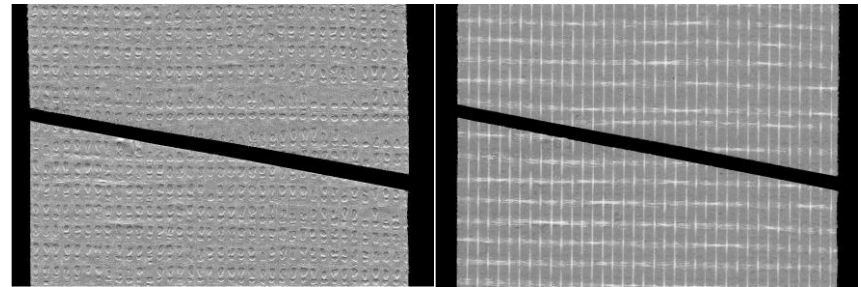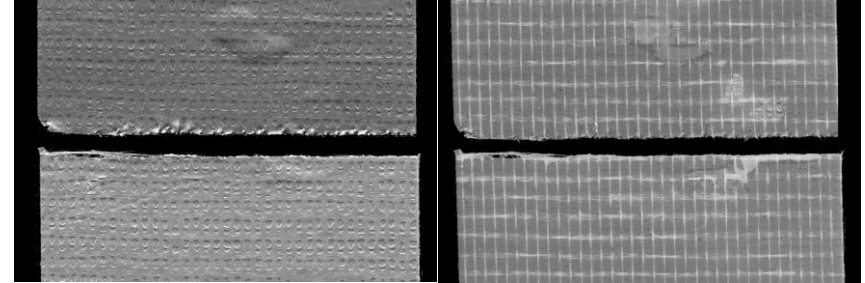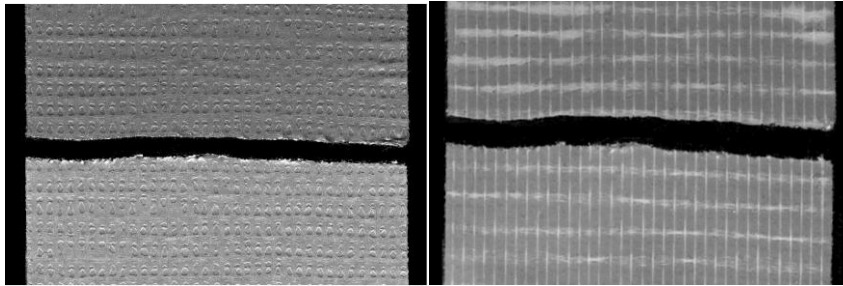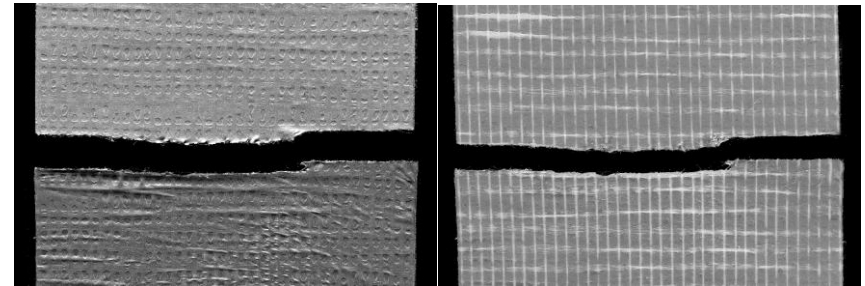Example 3.
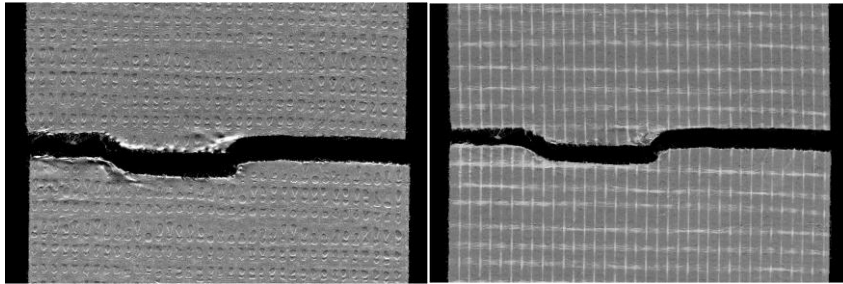
Example 4.
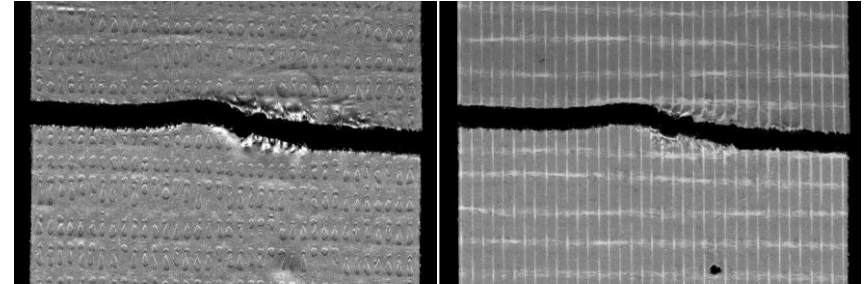
Example 5.

Example 6.

Figure 3.5. Examples of true matches from hand-torn set MQ-HT. Left of each pair: backing side, right of each pair: scrim adhesive side.

Examples 1 and 2 are of straight-angled tears, Examples 3 and 4 are wavy tears, and Examples 5 and 6 are puzzle-like.
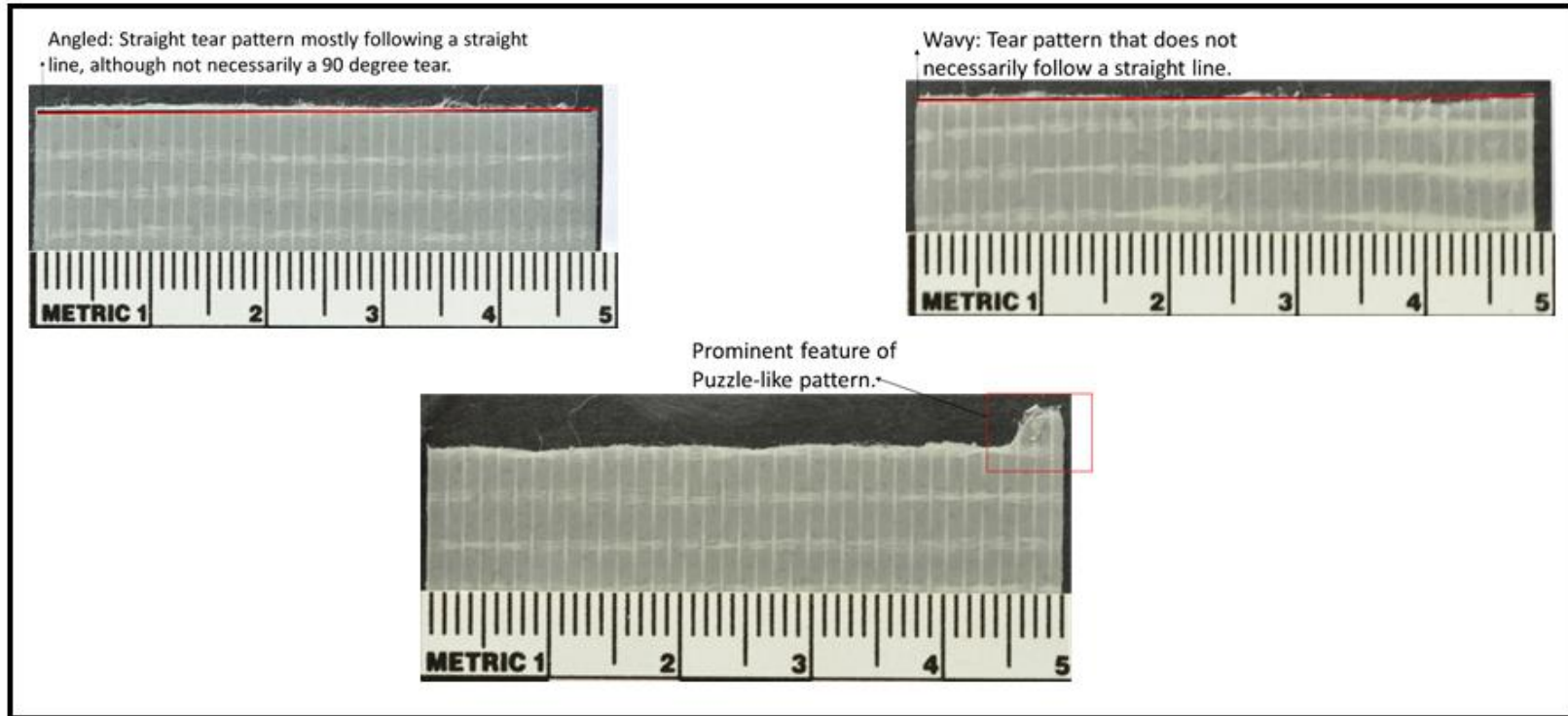
Figure 3.6. Descriptions of the three fracture patterns observed by the examiners on hand torn samples.

Table 3.3. Relative performance rates for set MQ-HT. The two analysts are described as Analyst A and Analyst B.

MQ-HT	Reported Non-Match	Reported Match	Reported Inconclusive	Total
True Non-Match	Analyst A and B: 409 (out of 409, 100%) **True Negatives**	Analyst A and B: 0 (out of 409, 0%) **False Positives**	Analyst A and B: 0 (out of 409, 0%)	409
True Match	A: 1 (out of 99, 1%) B: 2 (out of 99, 2%) **False Negatives**	Analyst A and B: 97 (out of 99, 98%) **True Positives**	A: 1 (out of 99, 1%) B: 0 (out of 99, 0%)	99

The accuracy rate for the MQ hand-torn set was at 99.6% for both examiners. The microscopic features of the puzzle-patterned showed higher scores (all true matches ranging from 90-100%), while the other two patterns had more varied scores (typically ranging from 80-100%). The edge patterns, particularly the puzzle-tears, as well as the visible scrim lines, allowed the examiners to correctly identify which tapes were true matches with a high degree of certainty. The true positives in this set had a wider distribution of scores compared to set MQ-SC, with the majority of positives resulting in a score of 80% or higher. Conversely, the true negative samples received scores 35% or below, and also more tapes were reported with scores in the range of 10-20% (Figure 3.7).

For this set, a score above 80 offered strong support for a match conclusion (M+), and a score below 20 offered strong support for a non-match conclusion (NM+). Weaker exclusions were observed for score values between 30-40% (NM-), while weaker associations were observed for scores between 70 to 85% (M-).

Although the edge patterns produced by a scissor cut were less distinctive than hand-torn separations, it was found that the scissor cut produced less stretching of the edges, which was also reflected by the higher thresholds for true negatives and true positives. The variability between individuals and between method of separation did not contribute to a substantial change in the overall accuracy of physical fit determinations. In addition, though the examiners had slightly different distributions for the ESS, the differences were smaller than 10% and so remained a valuable quantitative measure to inform their conclusion.
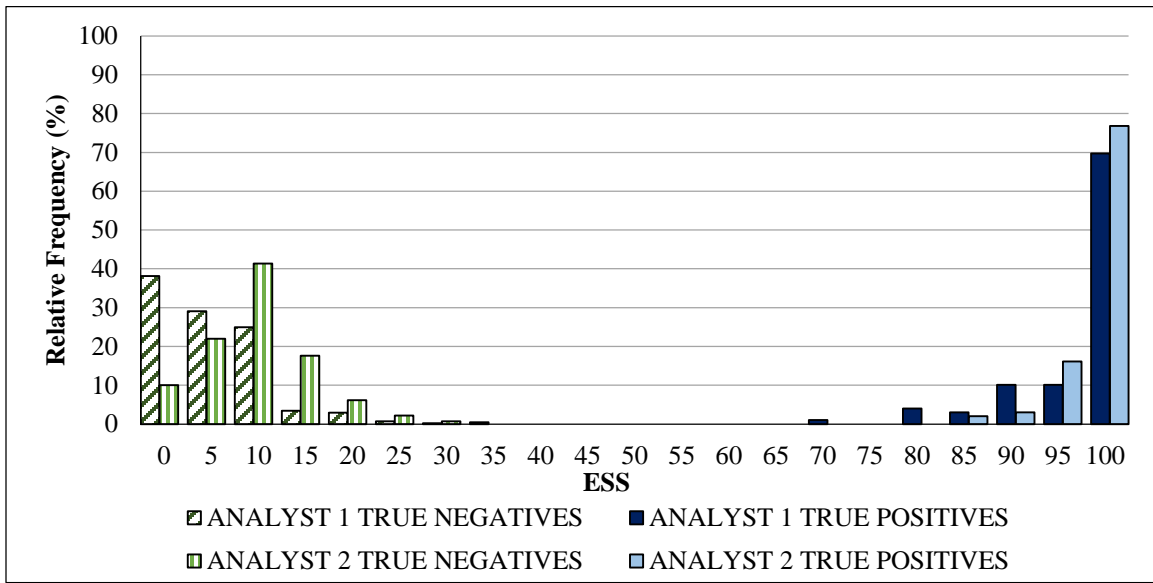
Figure 3.7. Relative frequencies of the occurrence of ESS for the 508 pairs evaluated for MQ-HT, by two examiners.

3.3.1.3. Hand Torn Stretched Tape Set – MQ-HT-S

Since stretching is an inherent concern in tape examinations, a subset of samples was moderately stretched to simulate a case process. Set MQ-HT-S was created by taking set MQ-HT and removing each tape from the acetate using liquid nitrogen. The tape was then pulled, stretched, and/or crumpled before it was smoothed back out and replaced in the same orientation. To avoid bias, the examiners participating in the comparisons were not the same individuals who performed the stretching. The examiner was not informed that the tape set was originally MQ-HT, the samples were relabeled, and the comparisons in MQ-HT-S were completed at least two months after the examiners had evaluated MQ-HT. Also, two of the tapes were removed from the set, leading to a total of 500 comparisons (403 non-matches, and 97 matches). Six example true matches are shown in **Error! Reference source not found.**.

Performance rates for the set are shown in Table 3.4. With this set, the only false negative was one of the tape pairs that had also been misclassified in MQ-HT.

Table 3.4. Relative performance rates for the hand-torn set MQ-HT-S.

MQ-HT-S	Reported Non-Match	Reported Match	Reported Inconclusive	Total
True Non-Match	403 (out of 403, 100%) **True Negatives**	0 (out of 403, 0%) **False Positives**	0 (out of 403, 0%)	403
True Match	1 (out of 97, 1%) **False Negatives**	96 (out of 97, 99%) **True Positives**	0 (out of 97, 0%)	97

The accuracy rate for this set was 99.6%. The overall edge patterns were the same as MQ-HT, with no tapes stretched in a way that would lead to a different pattern. The set had a comparable distribution of scores as MQ-HT, with the majority of positives resulting in a score 80% or higher and the majority of negative scores at 35% or below (**Error! Reference source not found.**). Like MQ-HT, a score above 80 offered strong support for a match conclusion, and a score below 20 offered strong support for a non-match conclusion. The stretching of the edges did contribute some degree to the variation seen in the score distributions and score-based likelihood ratios.

However, the degree of stretching applied and the effect of stretching on the edges did not affect the accuracy of physical fit determinations. There is the potential for more extensive stretching to shift the distributions of scores more than what was determined in MQ-HT-S, as discussed later in the mock case section.
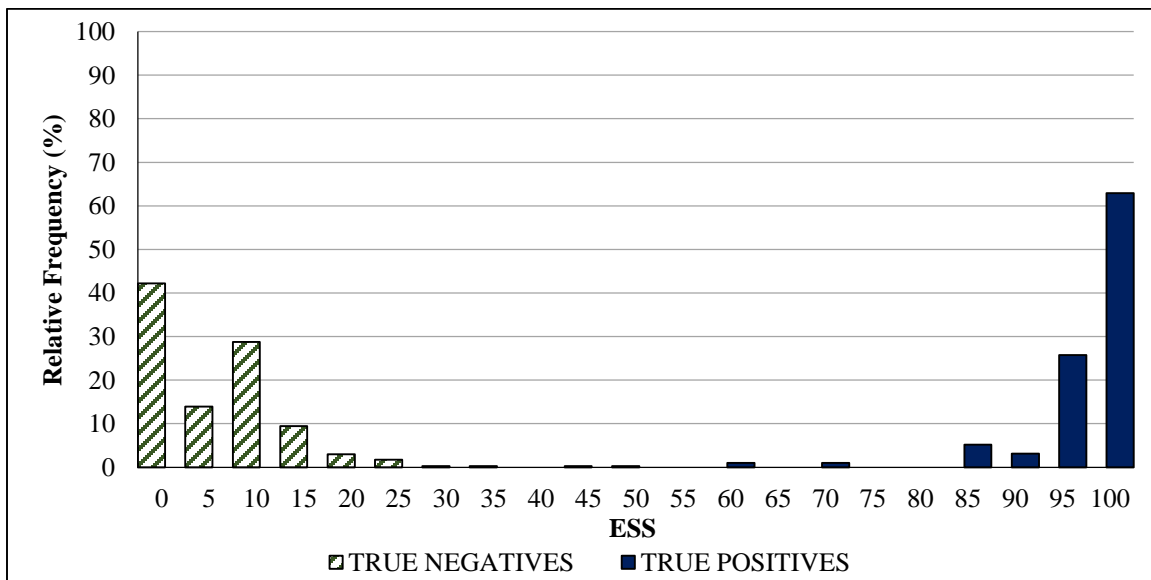
Figure 3.8. Relative frequencies of the match scores for the 500 comparisons for MQ-HT-S.

Example 1.
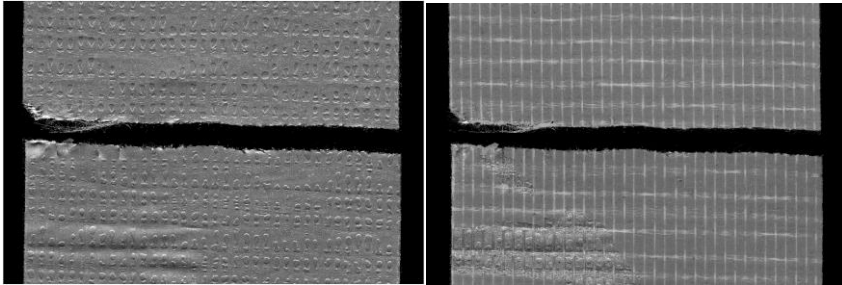
Example 2.
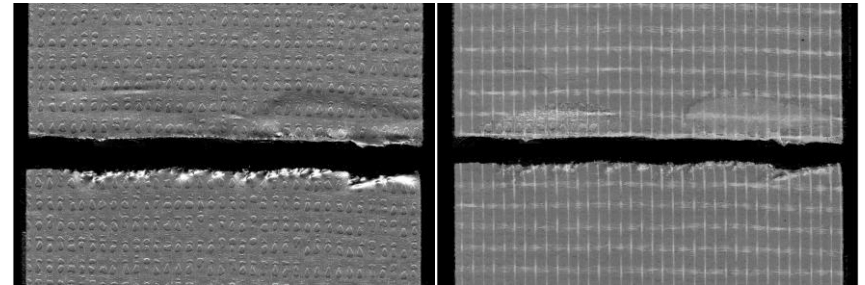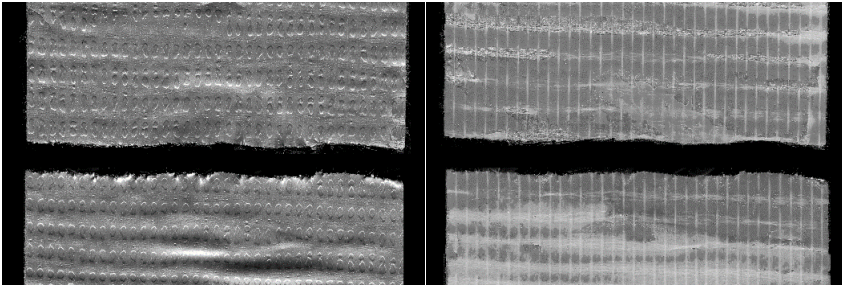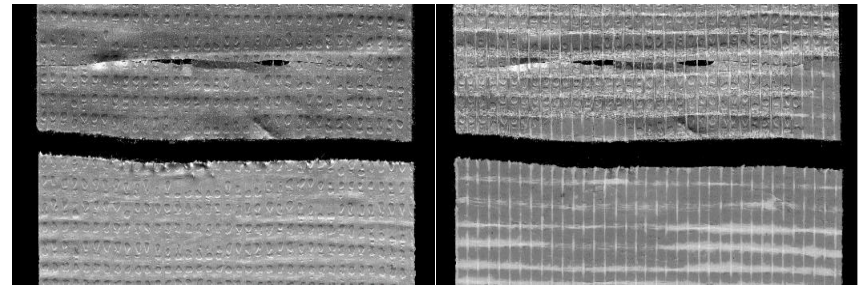
Example 3.
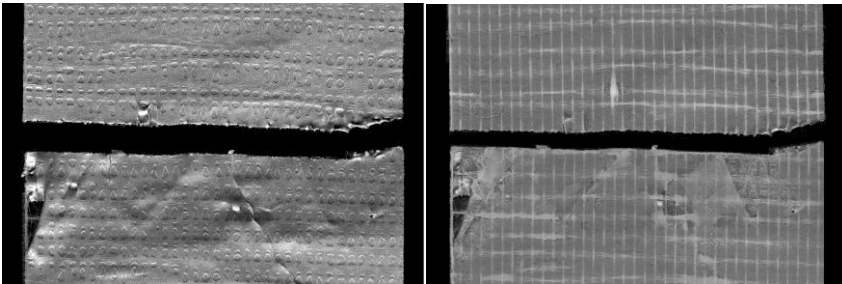
Example 4.
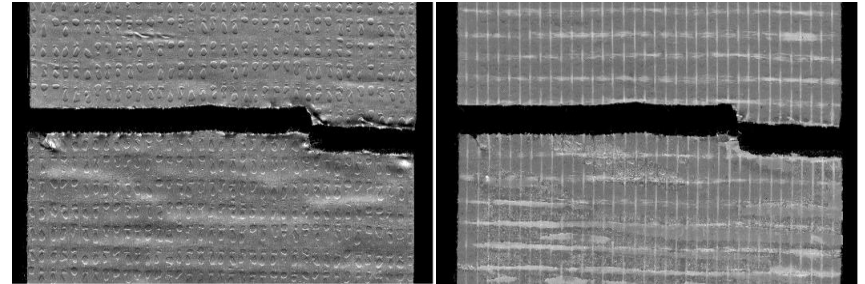
Example 5.

Example 6.

Figure 3.9. Examples of true matches from hand-torn set MQ-HT-S. Left of each pair: backing side, right of each pair: scrim adhesive side. Examples 1 and 2 are of straight-angled tears, Examples 3 and 4 are wavy tears, and Examples 5 and 6 are puzzle-like

3.3.2. High Quality Tape Set – HQ-HT

There were a number of challenges that arose from the high-quality set HQ-HT due to the thickness and cohesion of the adhesive. All the tapes in the HQ-HT set were torn in a single motion. The examiners removed the adhesive as they went, otherwise using the same conditions and microscopes as used for the mid-quality sets. It was noted during the preparation of the high-quality tapes that the tape tends to tear with an angled straight pattern, and only rarely left behind a puzzle-like edge pattern. No wavy patterns were obtained regardless of the tear motion. Examples are shown in Figure 3.10. A preliminary 85 pairs (referred to as HQ-HT-A) were completed to determine whether or not the comparison approach to the higher-quality samples needed to be adjusted. The set was found to have 4 tapes reported as false positives, a higher number of inconclusive and false negatives, and the distribution of the positive and negative scores had more variability than the mid-quality sets. True negatives generally received scores between 5-40% and the majority of positives received scores 75% or higher. The four false positives had very high ESS scores, over 90%. The score distributions for the subset are shown in Figure 3.11.

It was determined that introducing a transmitted light underneath the tape revealed microscopic features present at the edges of the HQ tear that were otherwise not evident. The light was not needed for the mid-quality sets, as when the light was on for those samples, the thinner backing and clear adhesive became transparent, and hid distinctive features necessary for comparison. In addition, if too much adhesive was removed on the edges, a higher degree of stretching and scrim shifting was noted, contributing to a high degree of false negative and inconclusive results. The microscopic tear features would have a corresponding direction had the tapes been a true match.

Instead, the indentations of the features mirror each other, indicating they are not a match. However, due to the relatively straight edges, those features were not evident without the transmitted source. An example is shown in Figure 3.12, illustrating both a true positive and a false positive edge with the respective directional tear features. These microscopic features are also visible on the underside of the tape, showing that the scrim areas do not correspond. It was also found that the straighter edges contributed to the uncertainty of the examiners.

After these factors were recognized, a set of 199 pairs (referred to as HQ-HT-B) were prepared to determine if the false positives seen in the first subset were a consistent aspect for the tape quality, or if introducing extra light and removing adhesive only from the areas at the edge could mitigate some of the uncertainty.

Error! Reference source not found. shows the results of the second subset. The additional precautions in the modified comparison protocol did eliminate additional false positives, although the occurrence of false negatives remained higher than other sets.

Table 3.5. Relative performance rates for the second set of HQ-HT tapes examined.

HQ-HT-B	Reported Non-Match	Reported Match	Reported Inconclusive	Total
True Non-Match	101 (out of 101, 100%) **True Negatives**	0 (out of 101, 0%) **False positives**	0 (out of 101, 0%)	101
True Match	21 (out of 98, 21.4%) **False negatives**	68 (out of 98, 69.4%) **True Positives**	9 (out of 98, 9.2%)	98

The accuracy of the subset HQ-HT-B was 84.9%. While the modified approach reduced the false positives, there were still a large number of false negative and inconclusive pairs. The examiners were more conservative in their conclusions for this set as a result of the need to remove the adhesive.

Although the examiners were careful, the potential for shifting and the straight tear patterns led to more uncertainty when comparing the edges. Despite the higher percentage of misclassifications, the distribution of scores for the true positives and true negatives for the set were similar to the distributions of the mid-quality sets (Figure 3.13). Weaker associations were observed for scores between 65 to 75%.

This set demonstrated that the quality of the tape could have a substantial effect on the outcomes of tape fracture identification. The uncertainty added when the adhesive must be removed can lead examiners to report more inconclusive and false negatives as a result of the alteration of some features. This set also demonstrated a need for examiners to be aware of additional considerations that would not be necessary for other grades of tape. However, the quantitative score remained a valuable predictor for the accuracy of the examination and to inform the examiner's opinion.
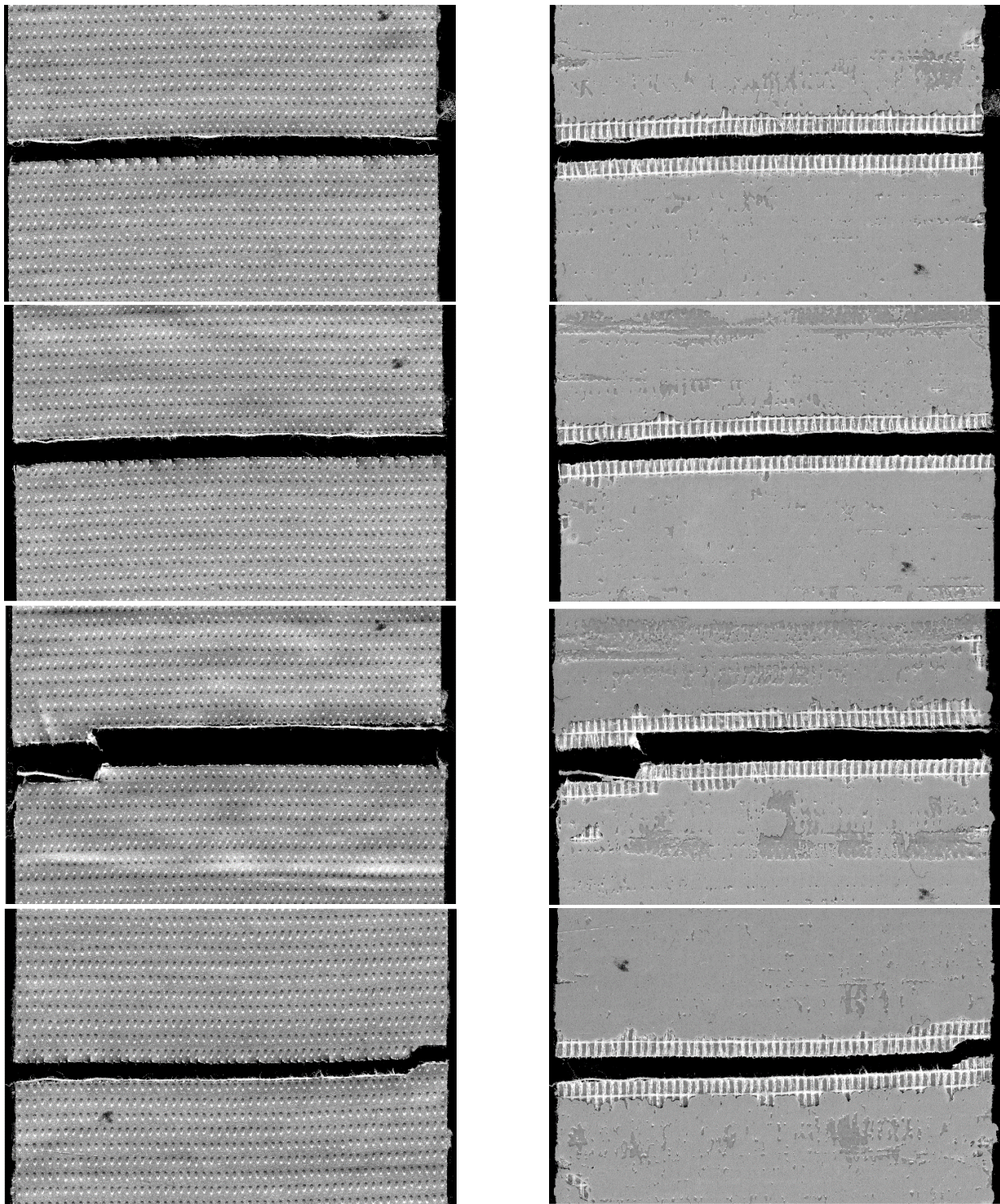
Figure 3.10. Examples of true matches from set HQ-HT. Backing sides are shown on the left and adhesive on the right. The set mainly had straight tears (as shown in the top two pairs), with a few with more distinct features (bottom two pairs).
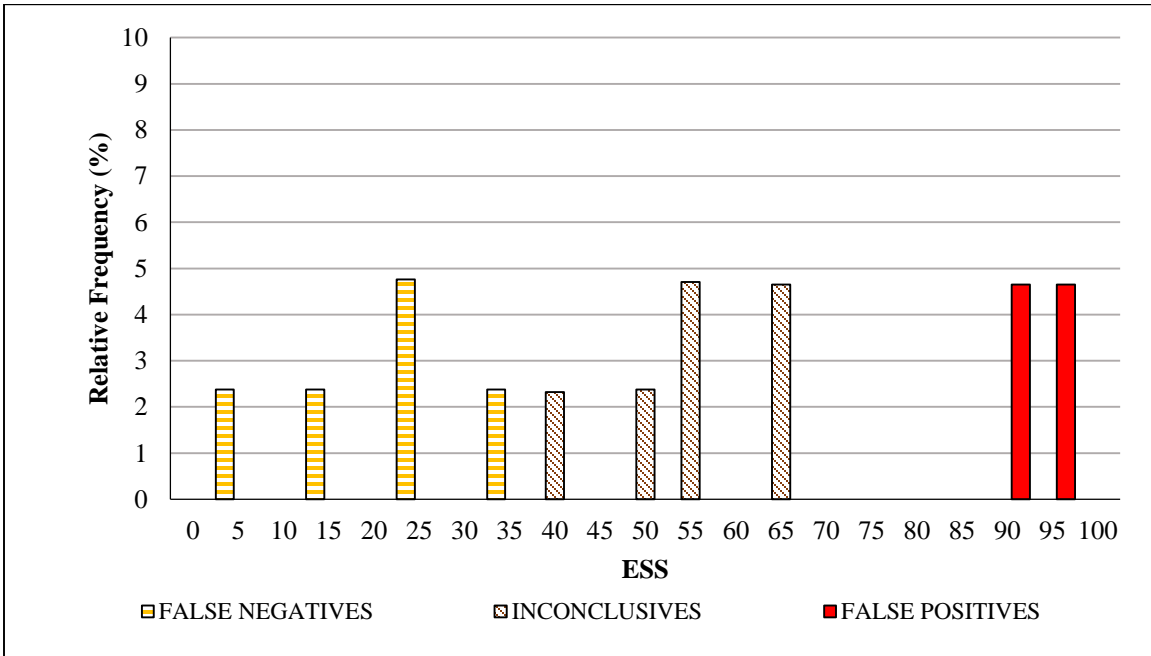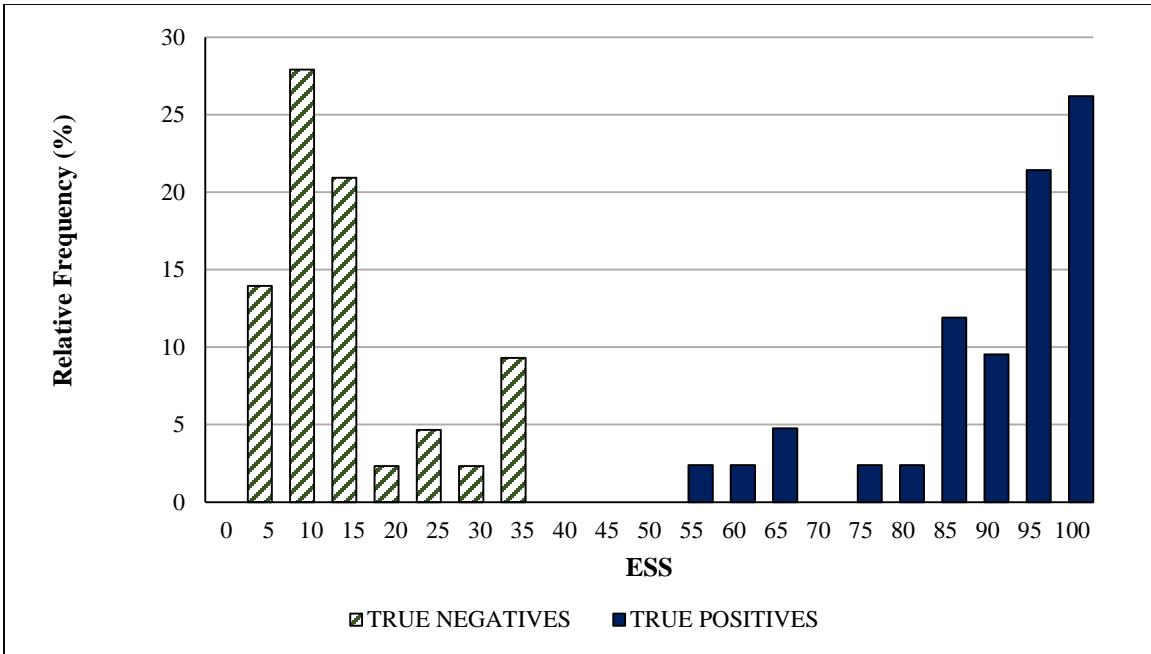
Figure 3.11. The top figure shows the distribution of the true positive and true negative scores for HQ-HT-A. The distribution of false negatives, inconclusive, and false positives are shown in the bottom figure.
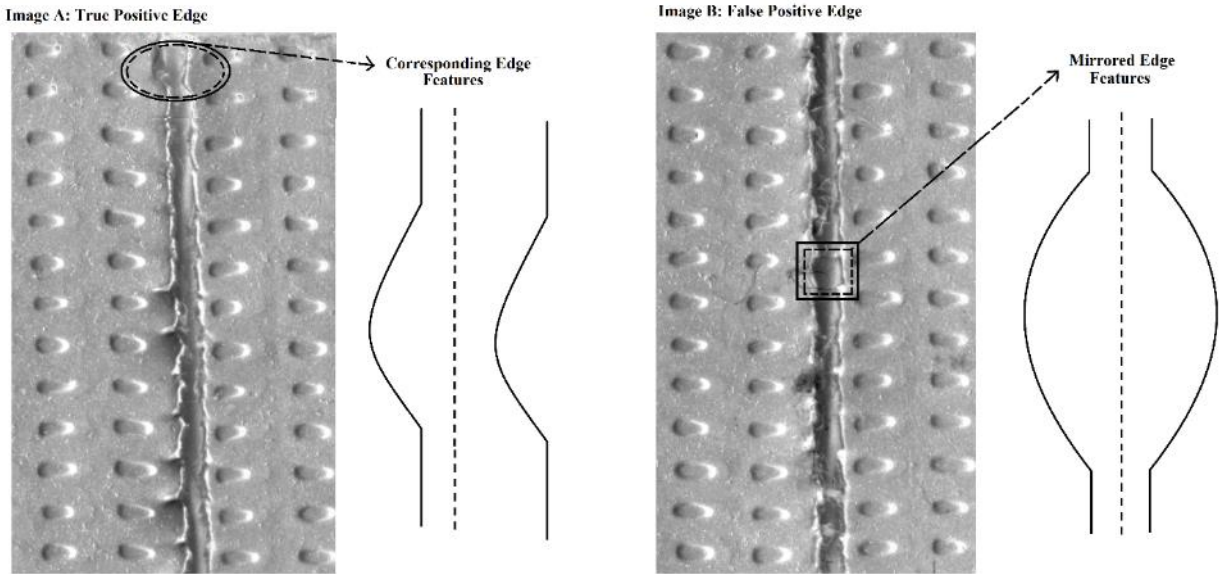
Figure 3.12. Example tears from HQ-HT. Images A shows a true positive edge. Circled in black are areas demonstrating corresponding edge features (as demonstrated with the example image on the right). Image B shows a false positive edge, with mirrored features highlighted in the black boxes.
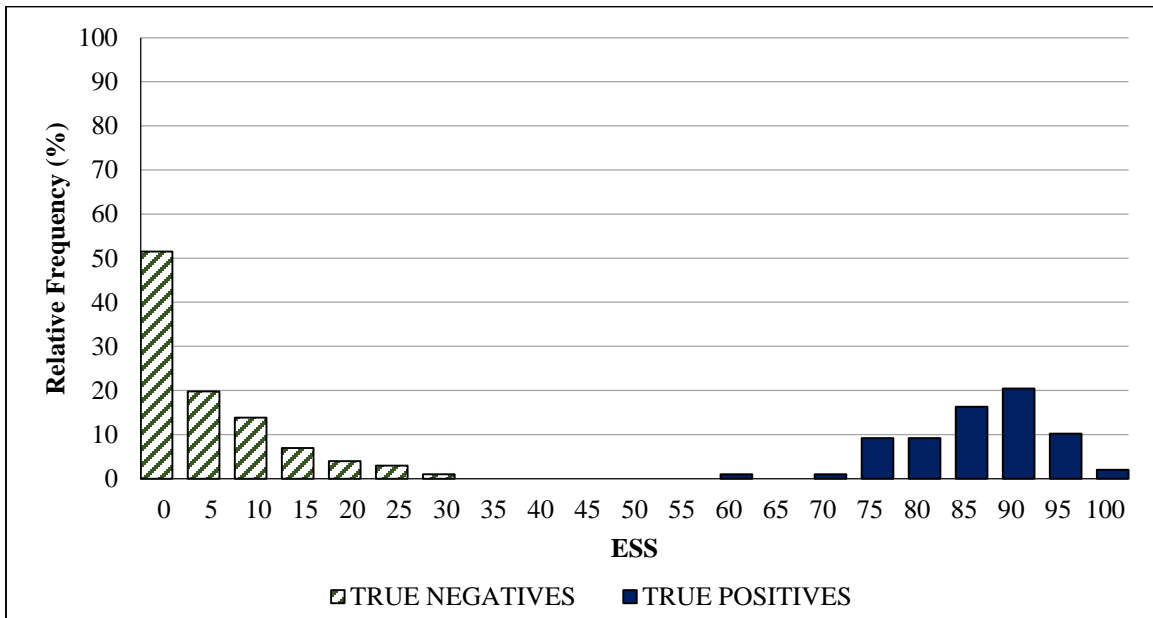
Figure 3.13. Relative frequencies of the match scores for the 199 comparisons in HQ-HT-B.

3.3.3. Low Quality Tape Set – LQ-HT

The low-quality set did not require the removal of the adhesive. Although not as transparent as the mid-quality sets, the adhesive was thinner, allowing observations of scrim features. It was determined during the preparation of the set that, unlike HQ-HT, the low-quality tape was far more likely to have a puzzle-like end. The thin backing and lower scrim count of the low-quality tape made it difficult to tear in a straight line, and so the majority of the tapes resulted in distorted features or edges with protrusions.

The set had a total of 200 comparisons, with 96 total true non-matches, and 104 true matches. Examples of true matches from the set are shown in **Error! Reference source not found.**. The performance rates are shown in Table 3.6. Only one true match pair was reported as inconclusive, while no false positives were observed. The overall accuracy for the set was 99.5%.

Table 3.6. Relative performance rates for LQ-HT.

LQ-HT	Reported Non-Match	Reported Match	Reported Inconclusive	Total
True Non-Match	95 (out of 96, 99%) **True Negatives**	0 (out of 96, 0%) **False Positives**	1 (out of 96, 1%)	96
True Match	0 (out of 104, 0%) **False Negatives**	104 (out of 104, 0%) **True Positives**	0 (out of 104, 0%)	104

Due to the distinct edges left after the tearing, the tapes that were true non-matches had many fewer areas of alignment than in other sets. On the contrary, the thin backing left some of the edges stretched, and as a result, the scores for the true matches resulted in less confidence in the match quality. **Error! Reference source not found.** shows the distribution of scores for LQ-HT. The distribution of scores for the true positives is more spread out, ranging from 65-100%. However, the majority of the true negative tapes had a score between 0-10%. The distribution of scores for the true negatives was substantially shifted towards zero.

Overall, the low-quality tapes lead to high certainty in tapes that were true non-matches. The true matches received scores in a wider range, demonstrating how the examiners had lower certainty in the quality of the matches they determined due to the distortion of the edges during the tearing.

The distribution of true positive scores was more similar to set HQ-HT. In both sets, the certainty of the match was lower than MQ, as a result of different characteristics of the tape.
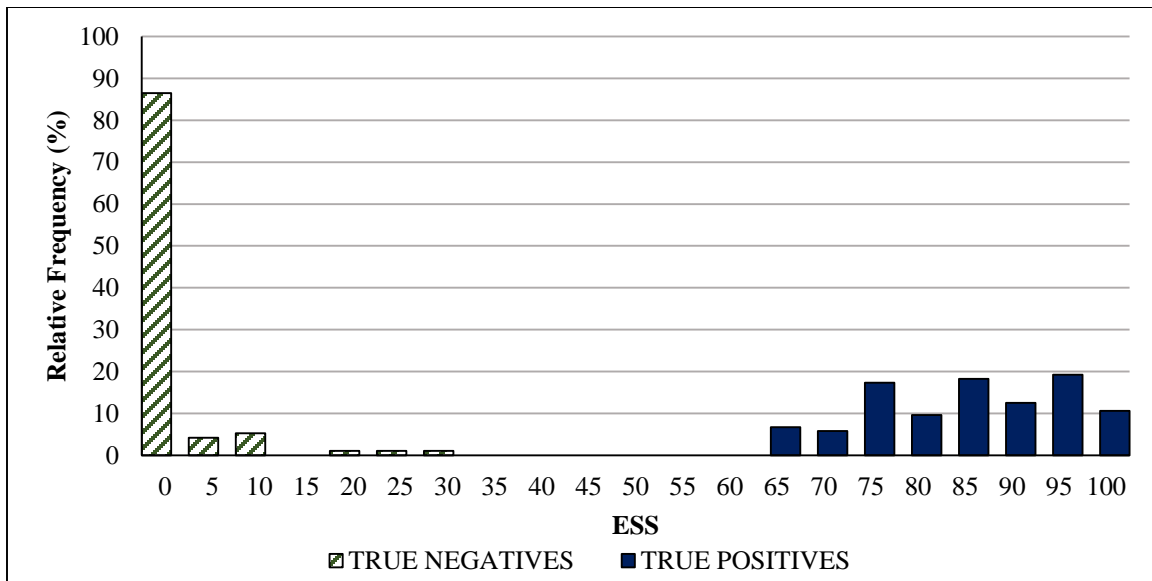
Figure 3.14. Distribution of true positive and true negative scores for LQ-HT.

3.3.4. Comparison of Tape Sets

The boxplots and the probability density functions (kernel density functions) were utilized to explore the trends in distribution and separation between true positives and true negatives by similarity score (Figure 3.16). The data demonstrates that there is a good separation between TP and TN distributions regardless of the factors considered. Overall, the mid-quality sets showed the highest degree of separation between the scores of true-matching tapes and true non-matching tapes. The scores for the TN tapes tended to fall between 0-35% for all sets, whereas the low and high-quality tapes resulted in a wider confidence range of values for the TP, as seen in the shift of the distributions to be between 60-100%. On the other hand, the mid-quality true positive distributions fell between 80-100%. A threshold of 40-60% ESS was indicative of inconclusive conclusions. In addition to the groups' separation, the boxplot also shows the consistency of ESS between examiners for the MQ-HT set.
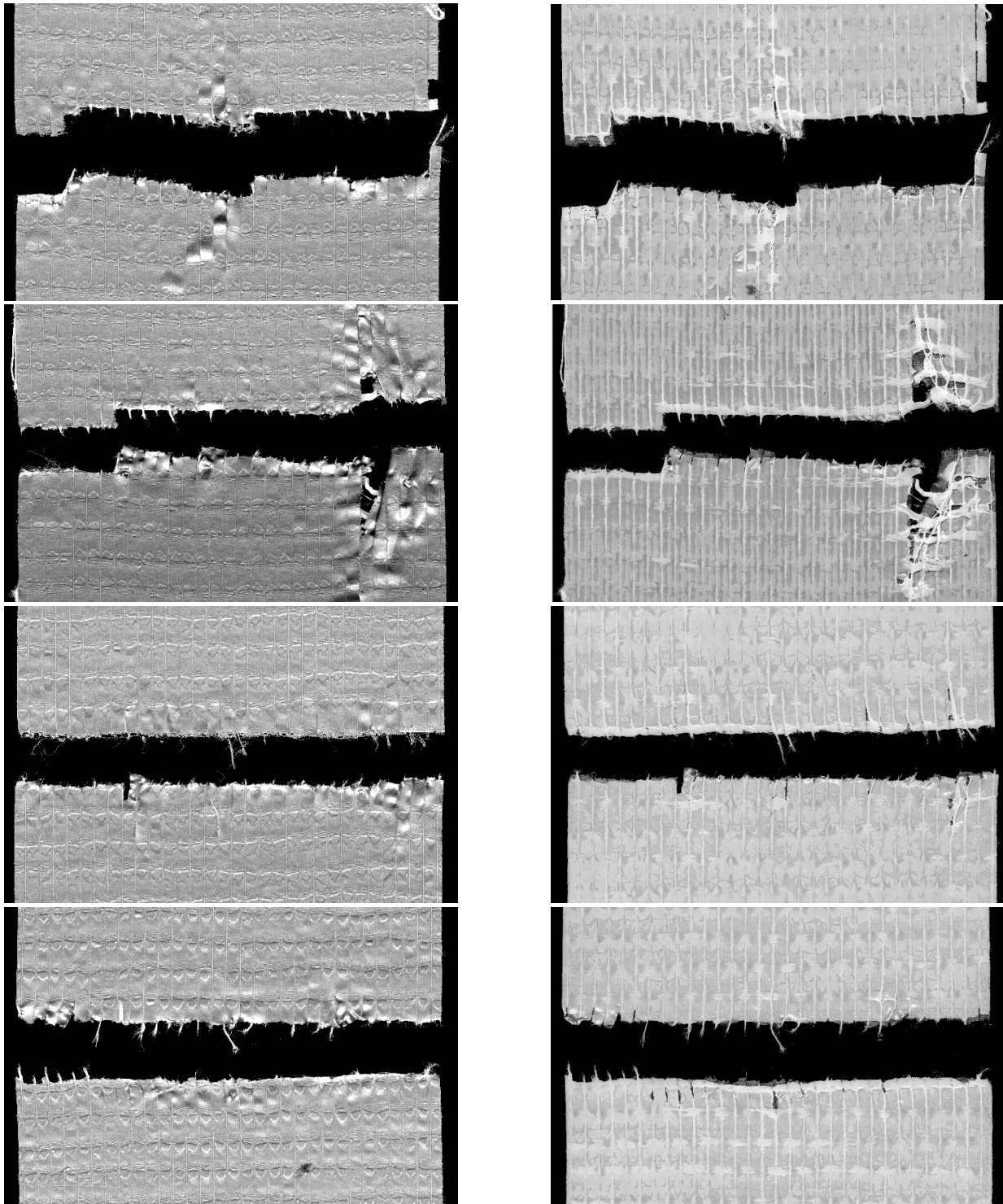
Figure 3.15. Examples of true matches from set LQ-HT. Left: backing side, right: scrim adhesive side. The set mainly had distinctive tears (as shown in the top two pairs), with a few with straighter edges (bottom two pairs).

ROC curves were generated from the data as an alternative method to illustrate the performance of the method in each set (Figure 3.17). The AUC was also calculated for each set. The performance differences between the three grades are shown in the curves, and in the AUC. High quality had the largest number of misidentifications, and correspondingly the lowest AUC. The plots for all three mid-quality sets are very close to ideal (following parallel to the y-axis then the x-axis), confirmed by AUC values very close to 1 (0.9925-0.9999). Even though some sets were imbalanced in terms of the number of true positives and true negatives samples, the performance of the method is in close agreement with results observed estimating performance rates from frequency data. In particular, the mid-quality sets had a ratio of 4 true non-matches to 1 true matches. To test the potential effect of the unbalanced dataset, the medium quality sets were re-evaluated by compiling a random subset of 100 true non-matches to include a 1:1 ratio of each class, and comparing the performance of those scores against the unbalanced full set. There was no significant difference on five random subsets compiled in R. The results could be an effect of the large separation that the TM and TNM had in these sets, but may not apply to other populations. Therefore, it is recommended to use a balanced set when possible, as designed for the high and low-quality sets.

To further explore the effect of the score distributions on the strength of the conclusions, score-based likelihood ratios were calculated. The quality of a duct tape severed edge is highly dependent on several factors, including the grade of the tape, the method of separation, degree of damage to the tape edges during collection as well as other influences. The factors may also be correlated, which further complicates the quality of the edge (i.e., a low-quality tape is more likely to be deformed by stretching).

Utilizing SLR values allows for a numeric score value to be translated into a probabilistic degree of certainty for the conclusion of match or non-match. Looking at an estimated distribution of SLR values can offer a practical insight into the weight of a particular match score reported during a physical fit comparison. For example, an ESS score of 75 translates into a high SLR value for a low or high-quality tape (SLR values of 2923 and 2711 respectively) whereas a score of 75 on a mid-quality tape only results in an SLR of 44. The SLR value still indicates support of a match

conclusion in a mid-quality tape. However, because the SLR values are dependent on the sample set, there must be some caution in relying purely on a calculated SLR when determining the value of the evidence.

The performance of a decision tool based on SLR can be measured by estimating misleading rates using Tippet plots. Tippet plots are cumulative versions of histograms where the proportion of cases is plotted versus the cumulative values of logarithmic SLR values for sets representing each of the alternative hypothesis (true match H_1, true non-match H_2). Tippet plots show the rates of misleading evidence at a given threshold, for example, where the value of the curve is at 0 (SLR = 1). Tippet plots can also be used to estimate discrimination power by examining the separation between curves. Figure 3.18 shows the probability distributions, the SLR frequency histograms, and the Tippet plots for all three grades of tape. The Tippet plots showed a distinct separation in the score-based likelihood ratios of same-source and different-source tapes. For the low and mid-qualities of tapes, there was a higher degree of separation at SLR = 1, indicating there were not many samples that were misleading in either set. On the contrary, the curve representing the same-source samples for the high quality did show the potential for a higher percentage of false negatives and inconclusive, which corresponded with the outcome of the identifications.

3.3.5. Case Study

A mock homicide case was designed at the crime scene complex to simulate a worse-case circumstance in which a victim has been bound with duct tape, and the samples were highly stretched in the crime and during the recovery of the evidence. There were a total of 18 samples collected from the crime scene: 6 known samples (12 edges) and 12 questioned samples (24 edges). The questioned tapes were labeled according to the location of the tape on the mannequin victim and which edge of the tape was removed (i.e., samples A1.1 and A1.2 were opposite ends of the same piece of tape). The examiners were only required to compare each questioned edge to the known samples for 288 comparisons total. Depending on whether the samples were adhered to acetate, plastic bags, or attached to the tape's backing or adhesive portions, different separation methods were attempted (cold, liquid nitrogen, solvent or heat). As the examiners were preparing the set, it was found that heat was not a suitable tool for this set, as the adhesive melted rather than softening, causing distortions in the backing, and the scrim to start shifting (Figure 3.19). For the remaining pieces, liquid nitrogen proved to be more effective.
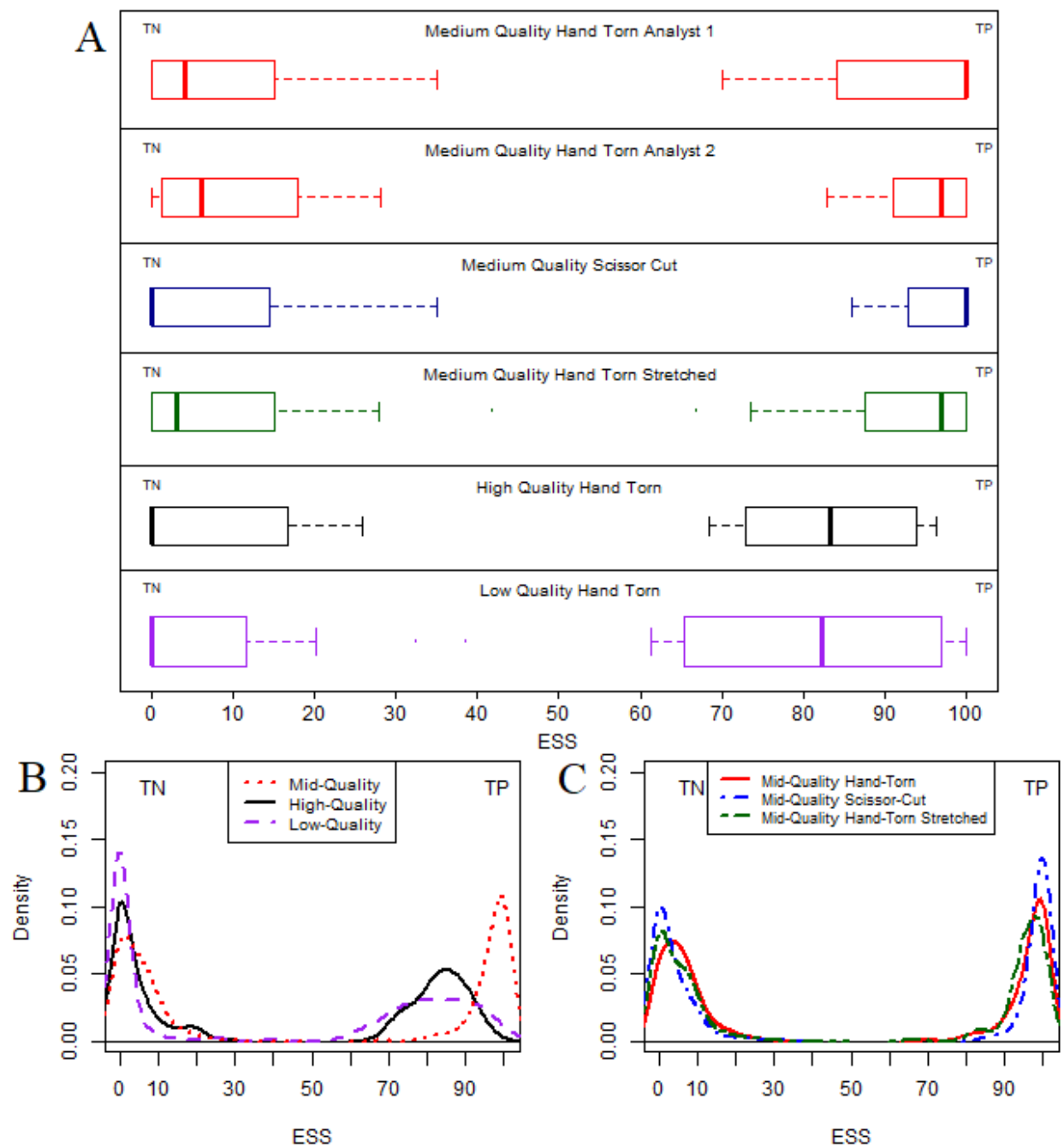
Figure 3.16. Image A shows the boxplots for all different tape sets considered in the study. Image B shows the probability density distributions for the three tape grades. Image C shows the breakdown of the tape sets that were mid-quality (pristine hand-torn, scissor-cut, and stretched hand-torn). The true negative distributions are on the left side of each figure, and the true positive distributions are shown on the right side.
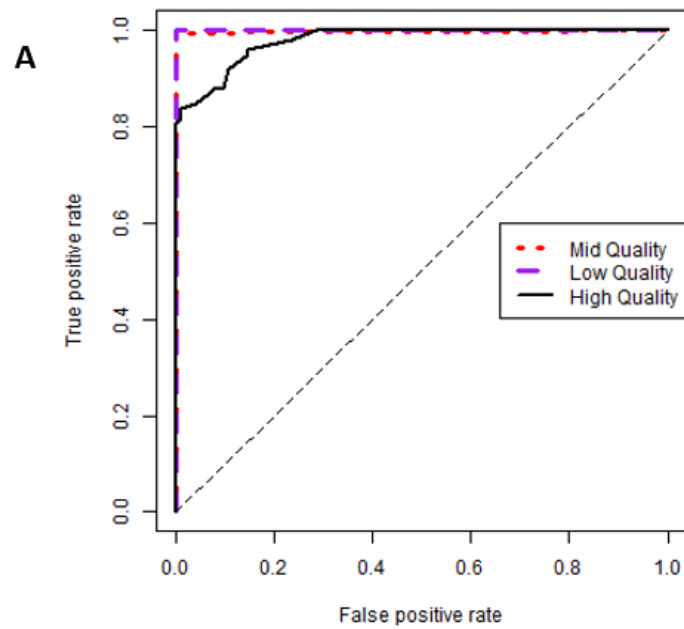
			Number of	
	Grade		Comparisons	AUC
	Mid		2016	0.99741
	High		199	0.97863
	Low		200	1

		Number of	
Set		Comparisons	AUC
Mid Hand Torn		1016	0.99885
Mid Scissor Cut		500	0.99254
Mid Stretched		500	0.99995
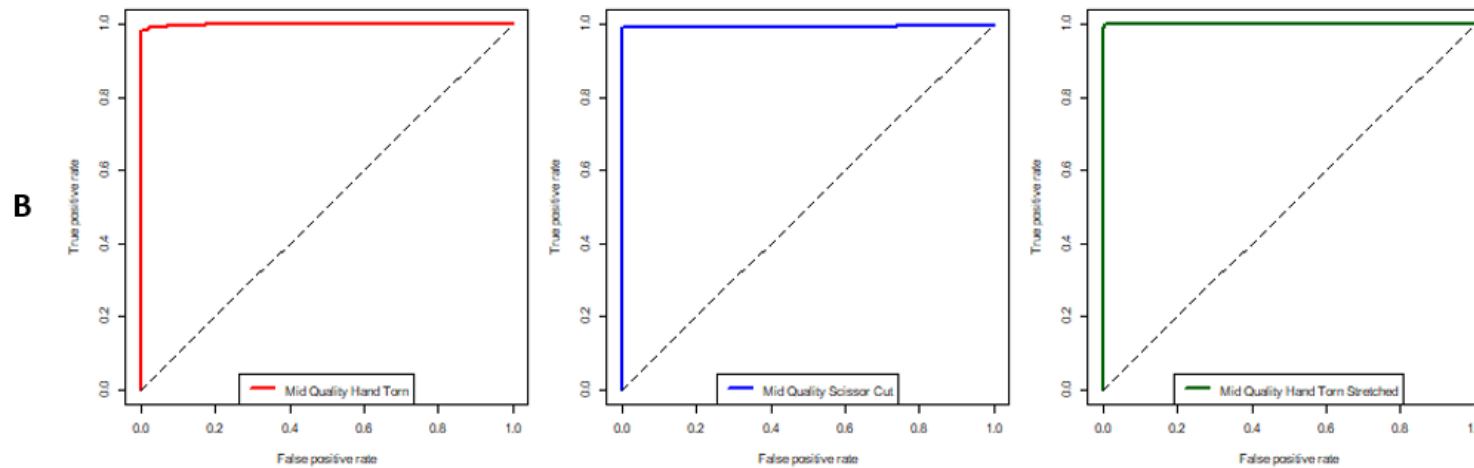
Figure 3.17. Image A compares the ROC curves for the three grades of tape. Image B shows the ROC curves for the three mid-quality sets. The AUC are shown for each set.
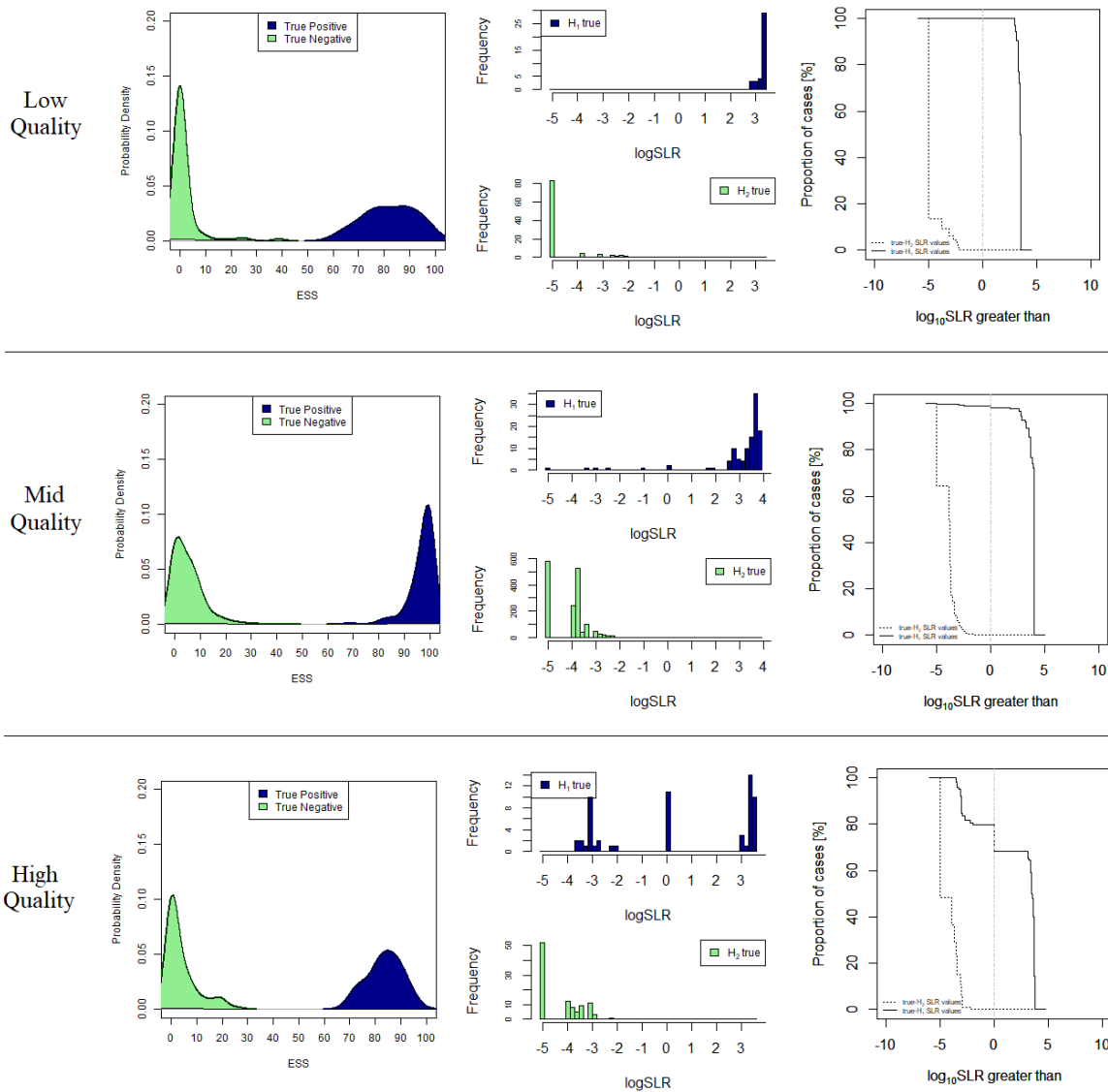
Figure 3.18. Plots representing the score probability density functions (left), frequency histograms of the logarithmic SLR values (center) and corresponding Tippet plots (right) for the low, mid, and high-quality tape sets.
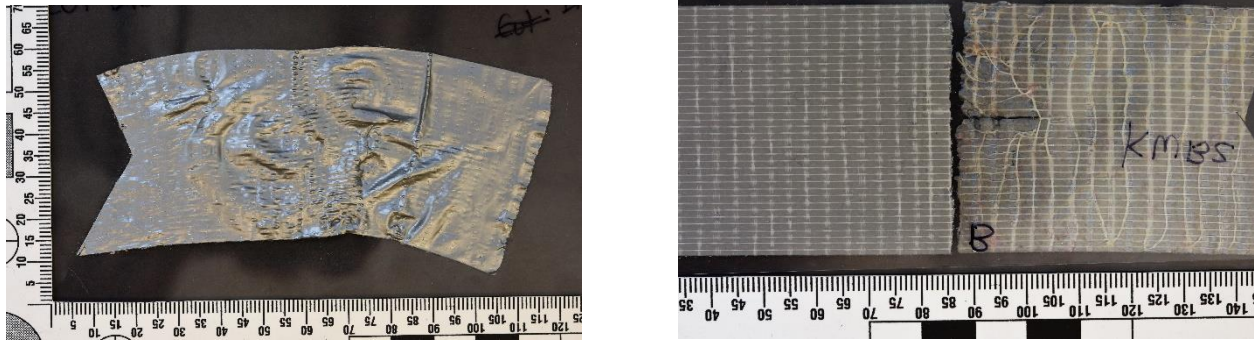
Figure 3.19. Case study tape samples prepared by the examiners. Image A shows a tape that was distorted after heating. Image B shows a distorted tape compared with the pristine known edge.

Two analysts independently examined the tape pairs. All of the non-matching tapes were correctly identified, with ESS ranging from 0-30%, and SLR ranging from 0.0001 to 0.01. No edges were incorrectly classified as a match. Of the nine total true matches, one examiner correctly identified six, and the other examiner identified five. The scores between the two examiners were averaged together, and the score-based likelihood ratio values were calculated as seen in Table 3.7. The tapes with high ESS received correspondingly high SLR values, indicating stronger support for the conclusion of a match. The tapes that received lower scores indicated support for the conclusion of a non-match, rendering three false negatives and one inconclusive.

While there were tapes misidentified as negatives from the case study, those samples would undergo the full chemical analysis with the known samples to characterize the composition and could be otherwise associated through the class characteristics. Moreover, regardless of the extreme stretching in this case study, there were 6 out of the 9 true matches that rendered high ESS scores and strong support via SLR.

Table 3.7. ESS results for true matches of tapes recovered from the mock crime scene.

Case Tape	Average ESS (n = 3)	Outcome	ESS Range (n=3)	SLR Range
D1.2	62	Inconclusive	61-70	20-40
E2.1	98	True Positive	91-100	2000-10000
E1.1	86	True Positive	81-90	400-2000
F2.1	90	True Positive	81-90	400-2000
E1.2	6	False Negative	1-10	0.0001-0.0003
F1.1	80	True Positive	71-80	40-400
A2.1	40	False Negative	31-40	0.02-0.09
A1.2	36	False Negative	31-40	0.02-0.09
C1.1	82	True Positive	81-90	400-2000

3.4. Conclusions- Duct Tapes Physical Fits

The goal of this study was to develop and validate a method for the quantitative assessment of the quality of a duct tape physical fit. Relative similarity scores computed per scrim bins served as a good predictor for fracture fit determinations. The examiners achieved high accuracy rates for the tape sets, with the low-quality grade at 99.5%, the mid-quality sets at 99.6-99.8%, and the high-quality set at 84.9%. False negative rates observed in the low and mid-quality sets under study were low (1-2%). Moderate stretching of the tapes did not have a substantial effect on the accuracy but did cause some variation in the score's distribution. Severe stretching represented by the mock case increased false negatives but still allowed true positives with a high degree of certainty. The variability between the method of separation (hand-torn vs. cut) had an insignificant effect on the overall accuracy of physical fit determinations.

When performing physical comparisons, examiners must be aware of factors that can either assist or hinder the determination of a physical fit. Each grade of tape has characteristics that may affect the comparisons. The high-quality tape used in this study introduced additional challenges and uncertainty to the classification, with a higher number of false negatives reported (21.4%). In the low-quality tapes, however, the performance rates were not affected despite the distortion of the edges.

This study demonstrated that not every true match holds the same certainty, and therefore, the basis of the reliability of physical matching are of relevance during a trace examination. The inclusion of score-based likelihood ratios applied to the ESS offers support for examiners performing tape classifications. The score-based likelihood ratio distributions allow for the

numeric value of a particular score to be translated into a degree of certainty for the conclusion of match or non-match. The SLR are intuitive and easy to assign; however, they are not capturing the rarity of features present on a trace, as a true likelihood ratio would. Alternative tools are being developed by other researchers to look at forensic evidence using computer-based models and applying other statistical frameworks to determine the weight of evidence [70-73] and are worth exploring in the future for the assessment of fracture fits.

While the overall matching observations are still a subjective decision for the examiner, the addition of quantitative descriptors of distinctive features can assist in the standardization of analytical protocols in the field. In this regard, a future study from our research group will assess the method as applied by multiple examiners experienced in physical fit determinations.

4. Assessment of the utility of X-ray Fluorescence for the chemical characterization and comparison of black electrical tape backings

NOTE: Parts of the following chapter, sections 4.1, 4.2, 4.3, and 4.5, are an adaptation of a previously published article ©2019: Prusinowski M, Mehltretter A, Martinez-Lopez C, Almirall J, Trejos T. Assessment of the utility of X-ray Fluorescence for the chemical characterization and comparison of black electrical tape backings. *For Chem*. 2019 (13). https://doi.org/10.1016/j.forc.2019.100146

A full copy of the published article is attached in Appendix 1.

4.1 Summary

Elemental analysis of electrical tapes is typically conducted by Scanning Electron Microscopy-Energy Dispersive Spectroscopy (SEM-EDS), although Laser Ablation-Inductively Coupled Plasma – Mass Spectrometry (LA-ICP-MS) recently showed to enhance the sensitivity of the determinations. In this study, the utility of X-ray Fluorescence (XRF) is compared to previously published SEM-EDS and LA-ICP-MS data. Three XRF systems were used to evaluate a range of configurations commonly available at crime laboratories. A set of 40 electrical tape backings known to originate from different sources was used to assess the inter-roll variability, discrimination and classification capabilities of the method. The discrimination for this tape set increased from 78.8% achieved by SEM-EDS to 81.5-91.0% by XRF, depending on the instrumental configuration. In comparison, LA-ICP-MS achieved 84.6% discrimination on these tapes. The overall characterization, classification, and discrimination capabilities for this set improved as follows: SEM-EDS<iXRF<small spot size benchtop XRF with SDD, LA-ICP-MS<large spot size XRF. A set of 20 pieces of tapes collected from the same roll were analyzed for intra-roll variability. Duplicate control same-source samples were used to evaluate inter-day and intra-day instrument variability. No false exclusions were observed in the data set, demonstrating the within-sample variability and instrumental variability are relatively lower than the inter-sample variability. One concern of the method is the penetration depth of the X-ray beam

beyond the target backing, requiring careful sample preparation to avoid interference from the adhesive or sample holder. XRF is a viable analytical tool for the forensic examination of electrical tapes, with advantages of speed of analysis, minimal destruction of the tape, and high informing power.

4.2. Methods and Materials

4.2.1 Instrumentation and measurement parameters

The XRF instruments used in this study included a μ-X-beam (X-beam slide, IXRF, Austin TX) coupled to an SEM-EDS (JSM-6490LV, JEOL, Tokyo Japan), a standalone XRF (ARL QUANT'X EDXRF, ThermoFisher, Waltham MA) and a standalone μ-XRF (M4 Tornado, Bruker, Billerica MA). The systems are referred to as A (IXRF), B (Thermo QuantX), and C (Bruker M4 Tornado), respectively.

Table 4.1 lists the acquisition parameters for the instrument systems. The data from system B were acquired using three different filters between the X-ray tube and the sample to optimize the signal-to-noise ratio of the species of interest. A 0.13 mm aluminum filter was used for lower energy source X-rays (Low Z), while a 0.06 mm palladium filter and a 0.559 mm thick copper filter were used for the mid-energy (Mid Z) and high-energy region (High Z), respectively.

4.2.2 Sample collection and sample preparation

A subset of 40 electrical tape backings was selected from a set of 90 previously characterized by our research groups [12,14] and measured on two of the XRF systems (A and B). Sample ID numbers are as described by Mehltretter et. al. [12] and Martinez et. al. [14] (samples 1 to 25, 27 to 41). A subset of 12 out of the 40 tapes was also analyzed on system C. A minimum of three replicates were collected from each sample.

The tapes were cut into ~1-2 cm^2 pieces, and the adhesive was removed from the backing using hexane. Three portions of each of the tape's backing were stored in pre-labeled plastic bags and coin envelopes.

Table 4.1. Instrumental parameters for the XRF instruments used in this study.

	A (IXRF-SEM-EDS)	B (Thermo QuantX XRF)	C (Bruker XRF)
Instrument	IXRF beam + JEOL JSM-6490LV + INCA x-sight EDS detector	Thermo ARL QUANT'X EDXRF	Bruker M4 Tornado XRF
X-ray source	Rh	Rh	Rh
Detector	SiLi (NCD)	SiLi (PCD)	Silicon Drift (SDD)
Spot size diameter	~150 µm	~1 cm	~25 µm
Voltage (kV)	50 kV	Low 12 kV, Mid 28 kV, High 50 kV	50 kV
Power (µA)	1000 µA	Low 600 µA, Mid 900 µA, High 1120 µA	500 µA
Working distance (mm)	17	54.1	10.5
Dead time (%)	12-18%	18-40%	~30%
Collection time (live seconds)	200	20	200

For samples measured using system A, the tape backings were coated with carbon (~100Å) and then mounted on a transparent XRF film (Chemplex, NY) on top of an SEM sample holder. To prevent interferences from the sample holder, the XRF film was suspended on an open space of the holder of about 1 cm diameter by 2 cm depth. Samples measured on system B were placed directly on top of the holder aperture (approximately 2 cm diameter aperture), with a beryllium window on top of the tape to reduce signal contributed by X-ray interaction with the chamber material. For instrument C, the samples were placed atop a graphite planchet. Differences in sample mounting were a result of the specific structure of chambers, sample stages, and holders.

A roll of electrical tape (Super 33+, Scotch 3M, Saint Paul MN) was selected to evaluate the intra-roll variability on instrument A. The entire roll was unwound, and 20 pieces of ~1-2 cm were removed every ~38 inches. The samples were prepared following the same procedure described above for instrument A.

Daily performance quality control tests used a copper/aluminum standard for calibration of the energy scale (x-axis). The calibration maintained the Cu Kα and Al Kα emission peaks within 5 eV of their literature values and monitored the peak resolution. Several tape samples were analyzed in duplicate as a daily monitor of instrumental variation over time. Duplicate controls, consisting of a piece of tape from the same roll, were analyzed as a daily monitor of instrument variation over time. Three replicate measurements were acquired for each duplicate control sample. The NIST standard reference material 1831, a soda-lime glass, was measured for instruments A and B every day as a QC control to monitor variability of the intensity of the iron and strontium peaks. Instrument C was monitored using manganese and zirconium for a daily control as directed by the manufacturer. Manganese is used as a standard element to determine the resolution of the detector. Zirconium is used during energy channel calibration of the detector because it falls approximately in the middle of the spectra.

4.2.3 Data analysis

The two approaches to comparing the data collected in this study were spectral overlay, and a semi-quantitative method in which the integrated areas under the peaks were calculated and

compared statistically utilizing Analysis of Variance (ANOVA) and Tukey-Kramer post-hoc test. Spectral overlay is commonly used in forensic spectrochemical analysis to quickly distinguish between samples of differing composition. For XRF analysis, spectral overlay allows the examiner to pre-screen the spectra and compare the elemental profile of the samples. The sample origin was kept blind to the analyst. Differences between two samples are based on presence versus absence of characteristic peaks, or differences in the relative heights and shapes of the peaks. The signal-to-noise ratios are a criterion used in identification of peaks that would be relevant for comparative analysis as shown by Ernst et. al. [65]. For a peak to be distinguishable, none of the replicate measurements should overlap any of the replicate measurements of the comparison sample. For example, this is the equivalent of a minimum-maximum range established for each relevant peak height. For the samples to be distinguishable, at least one element in the spectra needs to show no overlap in the peak range between the two samples. In this study, each pairwise comparison between tape samples was conducted with all the replicates to account for the variability within the sample [72].

A semi-quantitative approach can aid in distinguishing between samples, particularly those that were otherwise indistinguishable by spectral overlay. For this study, only data from samples collected on instrument A were used in the semi-quantitative assessment. The spectra were normalized prior to comparison to minimize the potential effects of tape thickness differences. The normalization involved dividing the spectra's counts at each energy interval by the sum of the overall counts of the individual spectra and multiplying by a factor of 100000. The integrated peak areas and the signal-to-noise ratios were calculated for each replicate spectra. The signal-to-noise ratios were used to determine which elements were above the limit of detection prior to statistical analysis [65]

Spectral comparisons were conducted using the Iridium Ultra Software (IXRF, version 1.4, Austin TX), Microsoft Excel (Excel 2016, Redmond WA), open-source R (version 3.5.0), and JMP (SAS Institute, JMP version 14.0.1, Cary NC). Spectral overlay, integrated peak areas and signal ranges, calculated signal-to-noise ratios of elements, and ANOVA were used for comparison of spectra as described by Ernst et. al. [65] and Martinez et. al. [14]. The measured ranges in peak positions are reported in Table 4.2.

Several element peaks were not able to utilize pre- or post-peak ranges because the element peaks overlapped with other peaks in that energy range. Instead, a larger range was used on one of the sides of the peak to calculate the background with the same number of channels as the analyte peak. Discrimination power (DP) was calculated as reported by Mehltretter et. al. [12] and Martinez et. al. [14].

Table 4.2. Energy ranges (keV) for select elements.

Elements	Prepeak	Peak	Postpeak
Al (Kα)	1.26-1.42	1.42-1.66	N.A
Si (Kα)	N.A	1.66-1.82	1.82-1.98
Cl (Kα)	2.31-2.53	2.53-2.75	N.A
Ca (Kα) + Sb (Lα)	3.24-3.49	3.49-3.74	N.A
Ba (Lα) + Ti (Kα)	4.31-4.41	4.41-4.62	4.62-4.72
Fe (Kα)	6.15-6.29	6.29-6.58	6.58-6.71
Zn (Kα)	8.36-8.50	8.50-8.78	8.78-8.92
Pb (Lα)	10.22-10.38	10.38-10.74	10.74-10.92
Br (Kα)	11.64-11.76	11.76-12.02	12.02-12.14
Cd (Kα)	22.32-22.68	22.68-23.42	23.42-23.78
Cr (Kα)	N.A	5.34-5.60	5.60-5.86
Mo (Kα)	17.16-17.29	17.29-17.56	17.56-17.69

4.3. Results and discussion

4.3.1 Evaluation of XRF for characterization and discrimination of tapes from different sources

4.3.1.1 Sample preparation considerations

Since the sensitivity of XRF is superior to SEM-EDS but lower than LA-ICP-MS methods, it was hypothesized that XRF would provide enhanced discrimination compared to SEM-EDS and a more cost-effective alternative to laser-ablation methods. With tape samples, the X-ray beam penetrates beyond the typical backing and adhesive thickness. The thickness of the electrical tape backings used in this study ranged from 96 to 198 μm, while XRF has the potential to penetrate up to several millimeters in depth depending on the atomic number of the element. As a result, samples should be prepared and exposed to the X-beam in such a way that prevents interference

on the X-ray spectrum from any material underneath the layer of interest. Unlike SEM-EDS and LA-ICP-MS, XRF requires removal of the adhesive layer from the backing, and the samples mounted on a substrate that do not produce X-rays artifacts. For comparative analysis, it is therefore critical that the known and questioned samples are prepared in the same manner. Moreover, the direct comparison of XRF data from one laboratory to another represents a challenge for XRF due to differences in backgrounds, sensitivities, and detection capabilities. Future development of standard reference materials for tape analysis may mitigate some of the limiting factors between instrumental configurations. At this point, forensic examiners should be aware of this limitation for sharing databases between laboratories.

For instrument A, the backings were suspended on X-ray film ~2 cm above the bottom of the sample holder. The diameter and depth of the holder were sufficiently large enough to prevent spectral interferences. Nonetheless, when the backing was sampled near the edges, artifact Cu peaks appeared in the spectrum from scattering of the holder walls. As a result, the areas of the spectra for Cu Kα and Cu Kβ (8.05 and 8.90 keV respectively) were omitted during the inter-sample and intra-sample comparisons.

The stand-alone XRF system B is configured to deliver the X-beam from underneath the sample holder and is equipped with a holder with a circular aperture of about 2 cm wide. The tape was directly attached to the aperture edges, and a beryllium disc was placed on top of the sample to reduce escape and scattering from the X-ray beam. Signal interference from the holder was negligible with this configuration.

Samples for system C were mounted on graphite planchets. Blank measurements of the planchets showed a peak at ~15.7 keV which was also observed in all sample spectra. Therefore, this peak was not considered in the comparisons.

4.3.1.2 Inter-source variability

A set of 40 tapes was selected from a previously characterized collection [8,12] purposely to include groups that show differences between various brands of tapes and groups that show similarity between same-brand tapes. The set consisted of tapes from 15 different brands and 30 different types of black electrical tapes. Table 4.3 summarizes the groups of indistinguishable

samples as measured on XRF instruments A and B. Using spectral overlay, SEM-EDS alone could classify the 40 samples (780 comparison pairs) into nine distinctive groups, while instrument A further discriminated the samples into 13 groups, and instrument B was able to classify samples into 19 groups from data acquired using all three filters. The spectral overlay conclusions were verified by an additional analyst prior to the finalized groupings.

The superior performance of the stand-alone system B is the result of the detection configuration, the beam size, and the use of filters [63]. The larger the area and thickness of the Si(Li) crystal, the better the efficiency to capture photons. Also, unlike other trace evidence, the typical size of tape specimens allows analysis from relatively large areas, which can be advantageous on systems with adjustable spot size. The beam size used in B was 1 cm diameter, allowing for a sampling area at least 4400 times greater than the micro-beam used on the other XRF instruments. Another benefit from the increased sensitivity is that the acquisition time can be reduced to only 20 seconds per replicate, speeding up the overall analysis time.

Inter-source variability, Instrument A

Figure 4.1 represents an example overlay of sample 21 shown in red and sample 38 shown in blue as measured by instrument A, previously not differentiated by SEM-EDS [12]. The elements found in the two samples include aluminum, chlorine, calcium, antimony, barium, titanium, iron and lead. Calcium, antimony, barium, and titanium were among the common elements found in the tapes measured and were oftentimes the most discriminating peaks between samples. However, standard resolution of EDS systems used here (0.1 keV), does not allow a baseline separation of the peak lines for antimony Lα (3.61 keV) and calcium Kα (3.69 keV) or the barium Lα (4.47 keV) and the titanium Kα (4.51 keV) peak lines. For samples 21 and 38, there was some degree of variation in intensity for the calcium and antimony peaks (~3.5-4.2 keV), but it was not sufficient to distinguish between the tape samples (Fig. 4.1).

Table 4.3. Samples grouped by SEM-EDS and XRF instruments A and B.

SEM – EDS Groups	Sample ID	XRF A Groups	Sample ID	Main differences by A	XRF B Groups	Sample ID	All Filters Main differences by B
1	4,8,32	1a	4		1i	4	Lower Ca than 1ii&1iii, higher Fe than 1ii
		1b	8,32	Higher Ca & higher Pb than 1a	1ii	8	
					1iii	32	Higher Ca than 1i&1ii
2	14,35,37	2	14,35,37		2	14,35,37	
3	21,38	3	21,38		3i	21	Higher Sb than 3ii
					3ii	38	Lower Ca, higher Fe than 3i
4	2, 10, 11, 12, 13, 15, 17, 18, 19, 20, 23, 24, 25, 27, 28, 39, 41	4a	2		4i	2	Lower Zn than 4ii&4iii
					4ii	10, 17, 23, 24	Lower Fe, higher Pb, higher Sb than 4iii
		4b	10, 11, 12, 13, 15, 17, 18, 19, 20, 23, 24, 25, 27, 28, 39, 41	Higher Zn than 4a	4iii	11, 12, 13, 15, 18, 20, 25, 27, 28, 39, 41	
					4iv	19	Lower Fe than 4i&4ii
5	16, 29, 30, 34, 36, 40	5	16, 29, 30, 34, 36, 40		5i	16, 29, 30, 34, 40	
					5ii	36	Contains Cr
6	1,5,7	6	1,5,7		6	1,5,7	
7	3,6,31	7a	3,6	Contains Ba/Ti	7i	3	
					7ii	6	Higher Ca than 7i, lower Ba than 7i&7iii, contains Cd
		7b	31	Higher Ca than 7a, no Ba/Ti	7iii	31	Higher Pb, higher Fe than 7i&7ii, no Ba
8	9, 33	8a	9		8i	9	
		8b	33	Lower Pb than 8a	8ii	33	Lower Pb, higher Fe, Zn than 8i
9	22	9	22		9	22	

Fig. 4.1. Spectra overlay comparison of tapes 21 (replicates 01-03) in red and 38 (replicates 01-03) in blue for instrument A.

Inter-source variability, Instrument B

The filters placed between the X-ray tube and the specimen significantly reduced the background of the spectrum measured by instrument B. Filters allow the transmission of X-Rays that are below the absorption edge energy of the filter's element, while absorbing other source X-rays. The result is a low background above the filter's absorption edge energy that enhances the signal to noise ratio in that region. Three different filters were used on instrument B to optimize the spectral region of the elements of interest. The aluminum filter was used for the low atomic number elements (Low Z) to measure elements with Kα lines between 2 and 10 KeV (e.g., Ca, Fe, Zn). The mid-range filter (Mid Z) made of thick palladium was used with a tube operating voltage of 28 KV, and improved detection of Br, Sr, Cr, Pb, and Mo. The thick copper filter was used with a tube operating at a high voltage (50 KV) for optimal detection of Mo, Ba, Pb, and Sb (High Z). Using the filters resulted in superior sensitivity and selectivity for the tape samples. For tapes 21 and 38, the superior signal-to-noise ratio allows for the two tapes to be differentiated by the amounts of calcium/antimony and iron (Fig. 4.2).
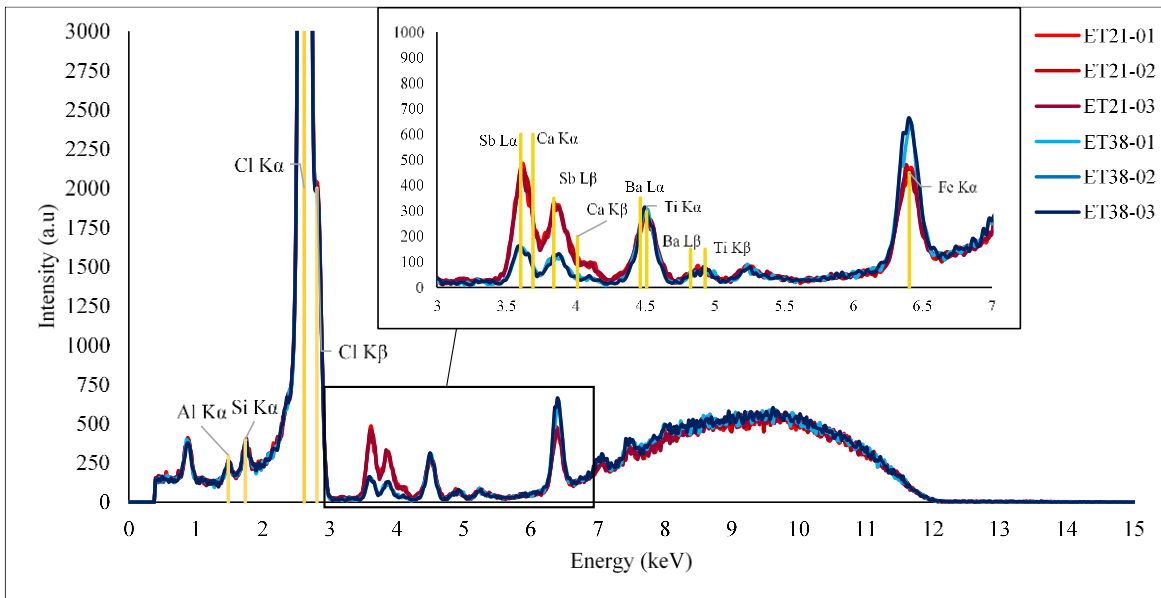
Fig. 4.2. Spectra overlay comparison of tapes 21 (replicates 01-03: red) and 38 (replicates 01-03: blue) for the low filter on instrument B.

In addition, the lack of resolution observed on EDS spectra for Ca and Sb peaks is clarified at the high Z, where Sb Kα and Kβ peaks, but not Ca, are present above 25 KeV. An example of this enhancement is shown in Figure 4.3 for the Sb peaks corresponding to samples 21 and 38. Likewise, the difficulties in resolving Ba from Ti peaks in the low KeV region are no longer problematic in the high energy region. The added selectivity represents a valuable advantage because these elements are highly informative and discriminating in electrical tapes.

4.3.1.2.1 Comparison of Instruments A and B for Inter-source variability

From the possible 780 possible comparison pairs from this set (n=40) as seen in Table 4.4, the XRF method B discriminated 90.1% when using combined spectral overlay results from all three filters. The discrimination capabilities were higher than SEM-EDS (78.8%), XRF attached to SEM (method A) (81.5%) and LA-ICP-MS (84.6%) for the set [12,14].
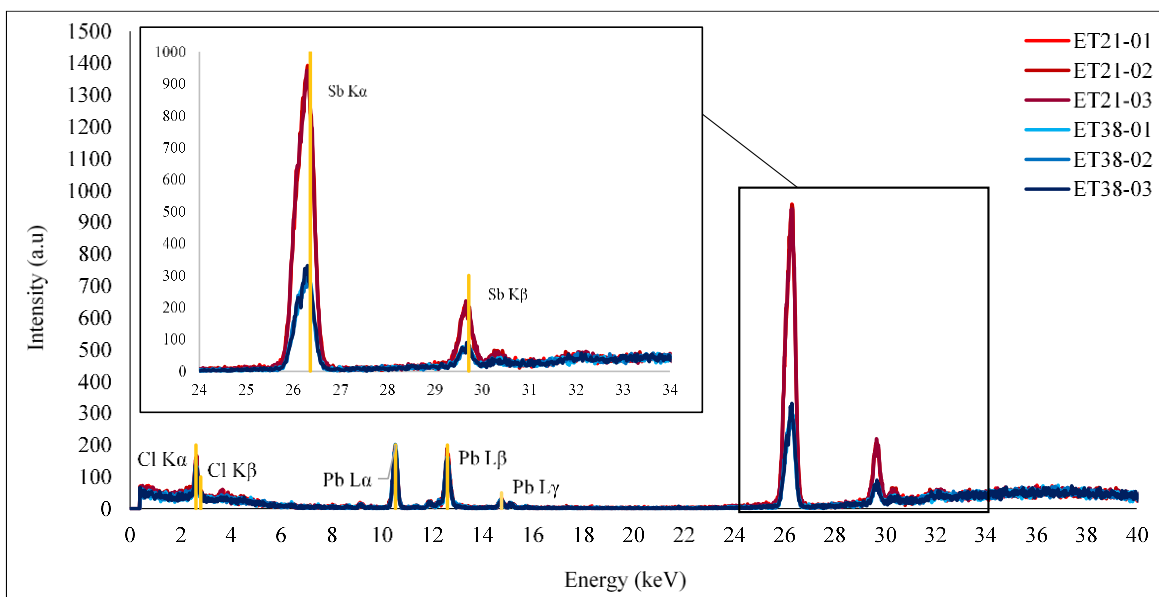
Fig. 4.3. Spectra overlay comparison of tapes 21 (replicates 01-03: red) and 38 (replicates 01-03: blue) for the high filter on instrument B.

Table 4.4. Summary of the discrimination capability for the SEM-EDS, XRF and LA-ICP-MS.

	SEM-EDS	XRF A	XRF B (low filter)	XRF B (mid filter)	XRF B (high filter)	XRF B (all filters)	LA-ICP-MS
Discrimination power (%)	78.8 (614 pairs out of 780)	81.5 (636 pairs out of 780)	90.0 (702 pairs out of 780)	81.8 (638 pairs out of 780)	86.1 (672 pairs out of 780)	90.1 (703 pairs out of 780)	84.6 (660 pairs out of 780)
Correct associations/ Controls (%)	N.A	100 (328 pairs)	100 (4 pairs)	100 (4 pairs)	100 (4 pairs)	100 (4 pairs)	100 (279 pairs)
Number of elements detected	8	14	10	9	8	14	29
# Distinct groups	9	13	18	16	17	19	19

The higher discrimination power of XRF instrument B compared to LA-ICP-MS was due to the capability of the instrument to classify the samples in group 4 (Table 3) into four separate classes based on the iron, lead and antimony content. Iron is more distinguishable using the aluminum filter, while antimony is resolved using the copper filter. Despite LA-ICP-MS having the capability

91

to detect more elements, the method was only able to separate tapes 2, 27 and 28 from the rest of the samples in that group. This is not surprising as iron detection is problematic in standard mass analyzers like the one used in this study [14]. Moreover, apart from tape 2, the tapes in that group originate from the same manufacturer, 3M [12]. Due to similar manufacturing processes, the composition of the tapes in that subgroup were very similar by XRF and LA-ICP-MS. Both XRF methods and LA-ICP-MS distinguished sample 2, which was manufactured by Advance®, while SEM-EDS did not detect the different compositions [12]. It is worth noting that, except for group 4, XRF and LA-ICP-MS methods classified the tapes into the same classes for this set of tapes. Nonetheless, LA-ICP-MS still provided superior overall selectivity and sensitivity that allows identification and characterization of a larger number of elements.

4.3.1.2.2 Comparison of Instruments A, B and C for Inter-source variability

A third XRF system was selected as an anticipated intermediate performance between A and B, and to represent another common XRF system found at crime laboratories. The second subset of 12 electrical tapes consisted of samples in 5 groups indistinguishable by SEM-EDS or XRF instrument A. Table 4.5 summarizes the spectral overlay group results as measured by the three XRF systems and the detected elements in each group.

Instruments B and C differentiated samples 21 and 38 by the calcium and antimony peaks as well as iron (Fig 4.4). However, unlike system B, instrument C was unable to resolve between the peaks of calcium and antimony in the energy range utilized (Fig 4.3).

XRF instruments A and C were unable to distinguish between samples originating from 3M (samples 10, 12, 15, and 24). On the other hand, instrument B could distinguish between the two brands of tape represented in the four samples (3M Super 33+ for samples 12 and 15, and 3M Super 88 for samples 10 and 24) based on the iron and lead content. Instrument B at high Z was also able to distinguish the samples by the antimony content using the higher energy range that was not measured on the other instruments (Fig 4.5).

Table 4.5. Samples grouped for the three XRF instruments.

XRF A Groups	Sample ID	Detected elements	XRF B Groups	Sample ID	Detected elements	Main differences by B	XRF C Groups	Sample ID	Detected elements	Main differences by C
1	8,32	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb	1i	8	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb, Br	Contains Br	1I	8	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb, Br	Contains Br
			1ii	32	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb	Higher Ca/Sb than 1i	1II	32	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb	Higher Ca/Sb than 1I
2	21,38	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb	2i	21	Al, Si, Cl, Ca, Ba/Ti, Fe, Zn, Pb, Br	Higher Sb than 2ii	2I	21	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb, Br	Higher Ca/Sb than 2II
			2ii	38	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb, Br	Higher Fe than 2i	2II	38	Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb, Br	
3	10,12,15, 24	Cl, Ca/Sb, Zn, Pb, Mo	3i	10, 24	Cl, Ca/Sb, Fe, Zn, Pb, Mo	Lower Fe, higher Pb and Sb than 3ii	3	10,12,15,24	Cl, Ca/Sb, Zn, Pb, Mo	
			3ii	12, 15	Cl, Ca/Sb, Fe, Zn, Pb, Mo					
4	16,36	Cl, Ca/Sb, Zn, Pb, Mo	4i	16	Cl, Ca/Sb, Fe, Zn, Pb, Mo		4I	16	Cl, Ca/Sb, Zn, Pb, Mo	
			4ii	36	Cl, Ca/Sb, Cr, Fe, Zn, Pb, Mo	Contains Cr	4II	36	Cl, Ca/Sb, Cr, Zn, Pb, Mo	Contains Cr
5	3,6	Cl, Ca/Sb, Ba/Ti, Pb	5i	3	Cl, Ca, Ba/Ti, Fe, Zn, Pb		5I	3	Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb	
			5ii	6	Cl, Ca, Ba/Ti, Fe, Zn, Pb, Cd	Higher Ca and lower Ba than 5i, contains Cd	5II	6	Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb	Higher Ca/Sb, higher Fe than 5I

Fig. 4.4. Spectra overlay comparison of tapes 21 (replicates 01-03: red) and 38 (replicates 01-03: blue) for instrument C.

Of the 66 possible comparison pairs for the twelve-tape subset, XRF method B had the highest discrimination power at 97% when using all three filters. Overall, XRF had high discrimination potential even for samples that were very similar in composition (Table 4.6).

Table 4.6. Summary of discriminating power for the XRF instruments in this study.

	XRF A	XRF B	XRF C	LA-ICP-MS
Samples from different sources	12 (66 pairs)	12 (66 pairs)	12 (66 pairs)	12 (66)
Discrimination power (%)	84.8 (56 pairs)	97.0 (64 pairs)	90.9 (60 pairs)	90.9 (60 pairs)
Number of Distinct groups	5	10	9	9

For the subset of 12 tapes, XRF system A performed comparably to SEM-EDS – the same tapes were indistinguishable for both instruments, resulting in a discrimination power of 84.8%. In contrast, system C and LA-ICP-MS separated these 12 tapes into the same groups, resulting in a discrimination power of 90.9%. While system B had the highest discrimination power of the three XRF instruments, instruments with this configuration are less commonly available in forensic laboratories as the other two.
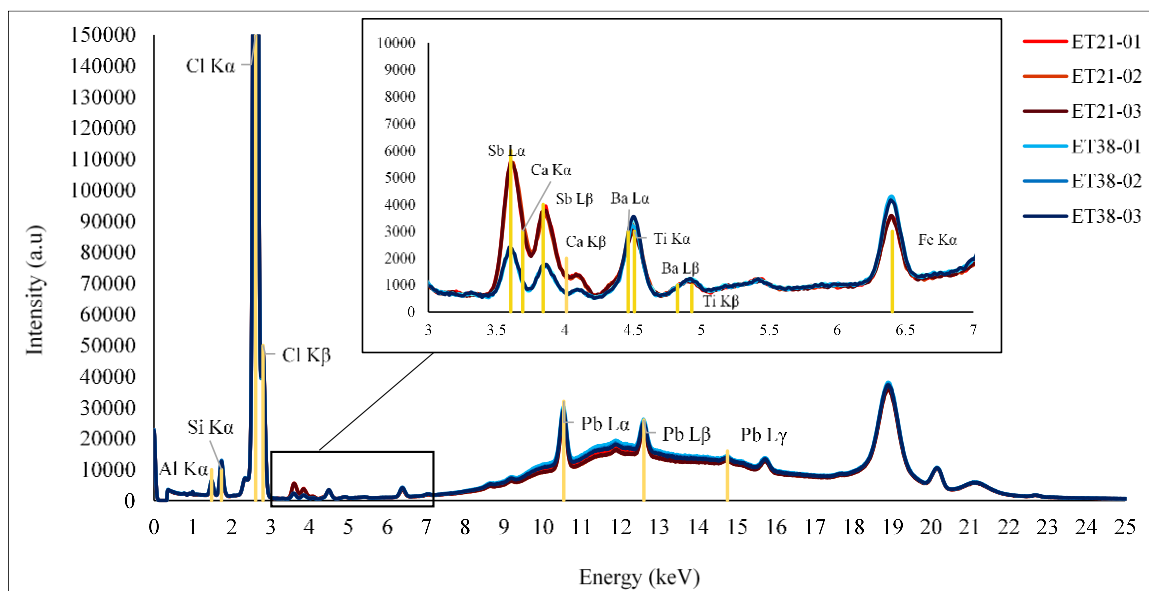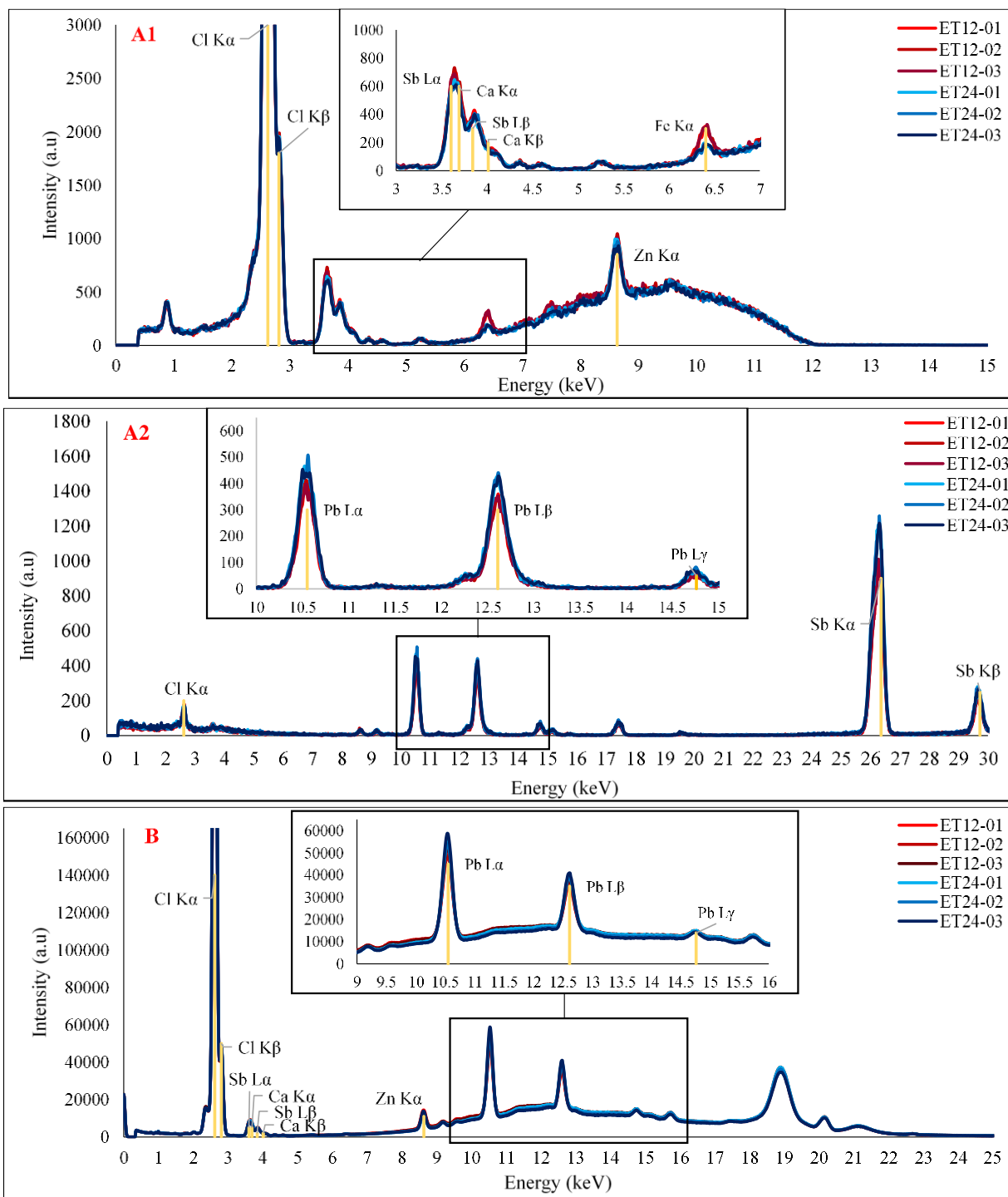
Fig. 4.5. Spectra overlay comparison of tapes 12 (replicates 01-03: red) and 24 (replicates 01-03: blue) for instrument B (image A1 for low filter and A2 for high filter) and instrument C (image B).

4.3.2 Example of characterization and semi-quantitative evaluation of sample variability using Instrument A

4.3.2.1 Comparison of intra-roll samples

ANOVA and the Tukey-Kramer test were used to determine which pairs of tapes had significant differences at each element. Quality control duplicates for sample 10 measured twice a day for each day of the study were analyzed to evaluate inter-day and intra-day variation for the instrument. Sample ET10 was collected as the first and last sample (duplicate controls 10A-10N) for each day of the inter-roll study. The ANOVA and Tukey-Kramer results for the comparison of the Cl/Ca ratio, calculated using the integrated peak areas, are shown in Figure 4.6. A graphical representation of the test can be depicted with circles, in which the center represents the mean of each group and the diameter is the HSD interval for that mean. If two circles overlap, then the two means are not different. The larger the overlap among circles, the more similar the groups are (larger p-values). If there is an angle of intersection less than 90 degrees between two circles, or no overlap in the circles, then the two samples are significantly different. The horizontal line represents the overall mean of all the samples, and the diamonds demonstrate the mean of each sample's replicate measurements. The measurements showed a low variation of the instrument over the course of the study. No significant differences in the means were found at $\alpha = 0.05$. The chlorine to calcium ratio was utilized to represent the worst-case scenario of within-sample variations, as it was observed by the spectral overlay comparisons that the peak from the polyvinyl chloride showed relatively poor repeatability.

In addition to the duplicate control samples, a set of 20 electrical tape backings (labeled ETV01-ETV20) was evaluated for elemental variability of samples known to originate from the same roll. The data for intra-roll variation were collected over the course of two days with sample ETV01 measured three times on the first day and six times on the second day to constrain inter-day and intra-day variation of the instrument. The signal-to-noise ratios of Cl, Ca/Sb, Fe, Zn, Pb, and Mo were calculated using the peak ranges defined in Table 4.2.

Fig 4.6. ANOVA and Tukey-Kramer analysis of the Cl/Ca ratio for sample 10 for the inter-day (A through G) and intra-day (1: morning, 2: afternoon) measurements, respectively.

The elements that were found to have signal-to-noise ratios higher than 3 for the roll were Cl, Ca/Sb, Zn, and Mo. The three replicates within each sample were checked for outliers using Grubb's test. One duplicate sample (ETV02) was found to have a replicate that was beyond the typical variation seen in these samples. With only two replicates remaining, it could not be included in the statistical analysis, and so was removed. The remaining 21 samples were compared with spectral overlay and additionally by ANOVA followed by Tukey-Kramer analysis for the selected elements (Fig. 4.7).

For the intra-roll study, no significant differences in the Cl/Ca ratio or in the intensities of Zn and Mo were determined. All samples were correctly associated with the others from the same roll using the combination of spectral overlay and quantitative analysis. Intra-roll variations were found to be lower than inter-roll variations for electrical tape samples. Figure 8 represents the Cl/Ca ratio and the signal intensities for elements Ca/Sb and Zn for samples classified into different groups on XRF A from the inter-roll study of the set of 12 similar tapes.
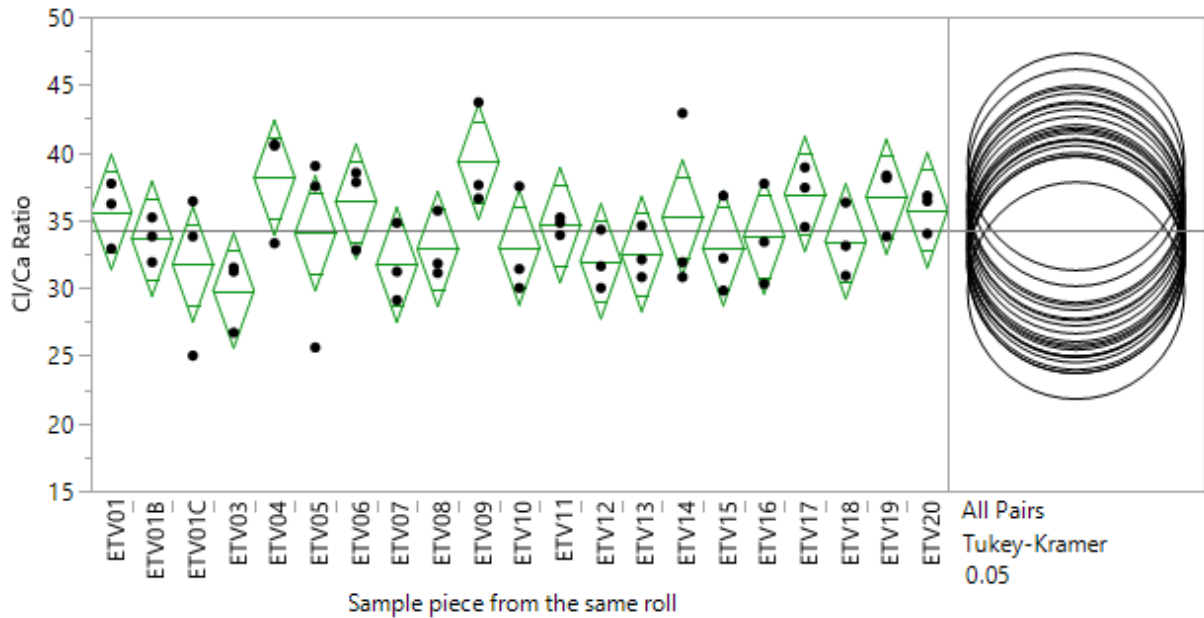
Fig 4.7. Analysis of the Cl/Ca ratio for the intra-roll samples by ANOVA and Tukey-Kramer.

As illustrated in Figure 4.8, samples 9 and 33 can be differentiated from all the other samples and each other utilizing the Cl/Ca ratio, which could not be accomplished with SEM-EDS [8]. Meanwhile, sample 31 can be differentiated by the amount of Ca and Sb, and sample 15 can be distinguished from most of the other samples by the Zn content. If one element is found to be significantly different between two samples, the tapes are found distinguishable. As such, relatively low within-sample and within-instrument variations are essential to recognize significant differences between samples from different sources.

4.4. Supplementary Data for XRF Analysis of Electrical Tape Backings on XRF A

A larger set of 89 electrical tape backings previously characterized by SEM-EDS and LA-ICP-MS was analyzed on XRF A [12,14]. The results of the characterization and classifications are displayed in Table 4.7.
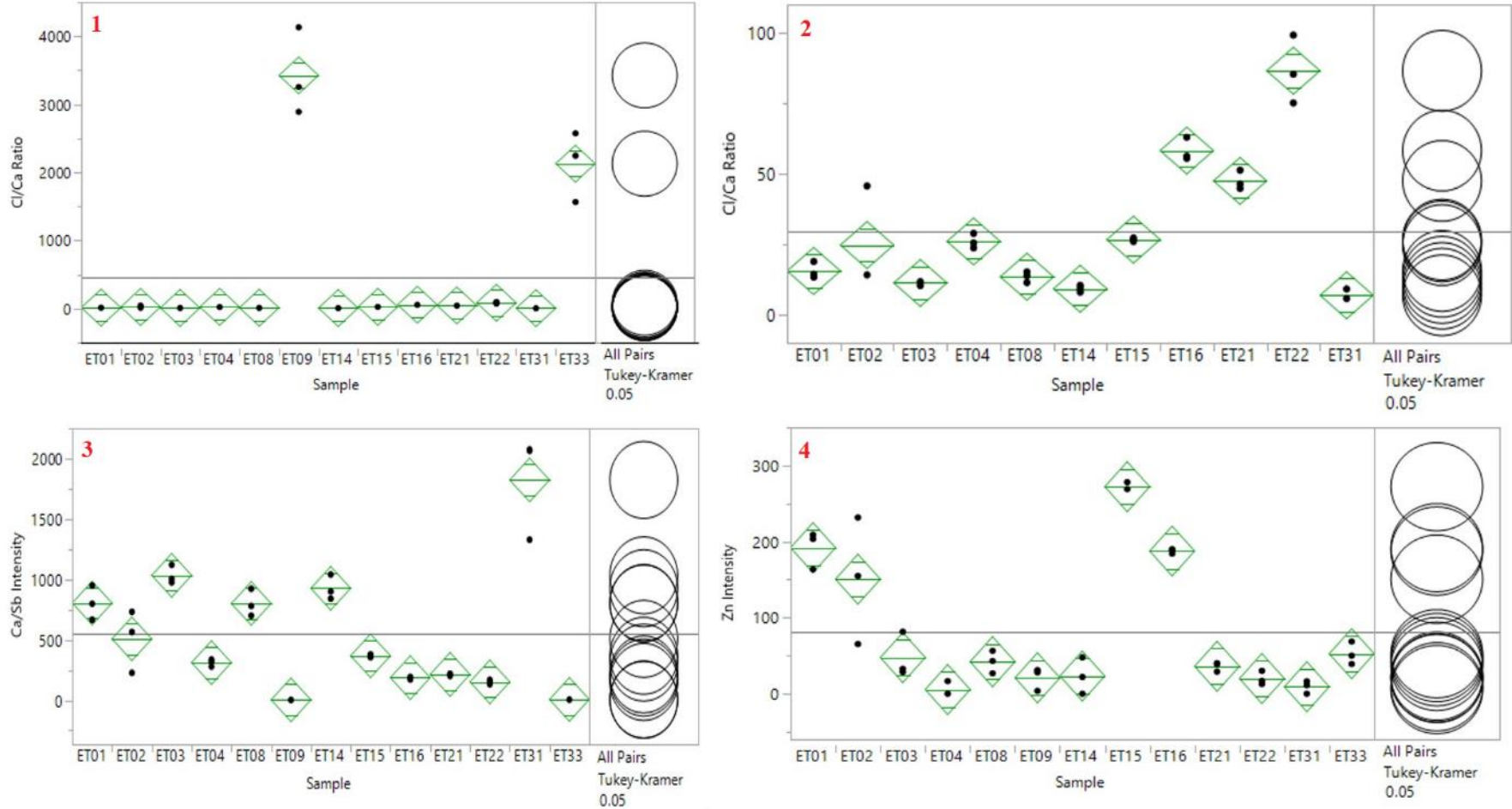
Fig 4.8. ANOVA and Tukey-Kramer analysis of tape samples from different sources. Image 1 demonstrates the Cl/Ca ratio analysis for some of the tape samples, while image 2 shows the variability of the Cl/Ca ratio for the same samples other than 9 and 33. Images 3 and 4 show the analysis of the signal intensities of Ca/Sb and Zn, respectively.

Table 4.7. Full set of samples grouped by SEM-EDS, XRF A and LA-ICP-MS.

SEM-EDS Groups	Sample ID	XRF Groups	Sample ID	LA-ICP-MS Groups	Sample ID
				i	4
		i	4	ii	42
				iii	45,55
				iv	51
				v	53
				vi	56
i	4, 8, 32, 42, 45, 51, 52, 53, 55, 56, 58, 70, 81, 82, 86	ii	8,32,52,81	vii	58
				viii	70
				ix	81
				x	82
				xi	86
		iii	42, 45, 51, 53, 55, 56, 58, 70, 82, 86	xii	8
				xiii	32
				xiv	52
ii	14, 35, 37, 50	iv	14,35,37,50	xv	14, 37
				xvi	35
				xvii	50
iii	21, 38, 46, 67	v	21,38,46,67	xviii	21,46
				xix	38
				xx	67
iv	66	vi	66	xxi	66
v	22, 69	vii	22, 69	xxii	22
				xxiii	69
vi	72, 74, 76, 77, 79, 80, 83	viii	72,74,79	xxiv	72
		ix	76,77,80,83	xxv	74,79
				xxvi	76,77,80,83
vii	62	x	62	xxvii	62
		xi	2,65	xxviii	2
viii	2, 10, 11, 12, 13, 15, 17, 18, 19, 20, 23, 24, 25, 27, 28, 39, 41, 54, 61, 63, 64, 65, 68	xii	10, 11, 12, 13, 15, 17, 18, 19, 20, 23, 24, 25, 27, 28, 39, 41, 54, 61, 64, 63, 68	xxix	10, 11, 12, 13, 15, 17, 18, 19, 20, 23, 24, 25, 39, 41, 54, 61, 63, 64, 68
				xxx	65
				xxxi	27,28
ix	16, 29, 30, 34, 36, 40, 43, 44, 47	xiii	16, 29, 30, 34, 36, 40, 43, 44, 47	xxxii	16, 29, 30, 34, 40, 43, 44, 47
				xxxiii	36
x	1, 5, 7, 48, 49, 57, 78, 84	xiv	1,5,7, 48, 49, 57, 84	xxxiv	1, 5, 7, 48, 49, 57
		xv	78	xxxv	78
				xxxvi	84
		xvi	3,6, 87	xxxvii	3
		xvii	31,71, 89,90	xxxviii	6
				xxxix	31
xi	3, 6, 31, 71, 87, 88, 89, 90			xl	71
				xli	87
		xviii	88	xlii	88
				xliii	89
				xliv	90
xii	73, 85	xix	73	xlv	73
		xx	85	xlvi	85
xiii	9, 33	xxi	9	xlvii	9
		xxii	33	xlviii	33
xiv	59, 60	xxiii	59, 60	xlix	59,60
xv	75	xxiv	75	l	75

Of the possible 3916 possible comparison pairs from this set (n=89), XRF discriminated 91.0%, as seen in Table 4.8. The discrimination capabilities were higher than SEM-EDS (87.6%) and lower than LA-ICP-MS at 93.0% [12,14]. The enhancement in discrimination seems to be directly correlated with the superior sensitivity (SEM-EDS<XRF<LA-ICP-MS), For instance, SEM-EDS is often limited to the detection of elements presents at concentrations greater than 1000ppm, while XRF can detect in the range of 30-100ppm, depending on the atomic number and the instrument configuration. LA-ICP-MS can typically reach detection limits in the single-digit ppm to sub-ppm levels. As a result, the number of relevant elements detectable by these methods increased from 8 by SEM-EDS, 14 by XRF and 29 by LA-ICP-MS. The superior performance of the method is also related to the selectivity of the techniques. For instance, although SEM-EDS and XRF suffer from similar resolution issues for low energy emission lines of Ca/Sb and Ba/Ti, the higher energy beam and penetration depth of X-rays allow the detection of alternative x-ray limes at higher energies (20-30 KeV region) that are well resolved. This was particularly important for the identification of Sb and Ba, which have shown to be highly discriminating in this dataset.

Table 4.8. Summary of the discrimination capability for SEM-EDS, XRF and LA-ICP-MS.

	SEM-EDS	XRF A	LA-ICP-MS
Samples from different sources	89 (3916 pairs)	89 (3916 pairs)	89 (3916 pairs)
Discrimination power (%)	87.6 (3429 pairs)	91.0 (3565 pairs)	94.2 (3690 pairs)
Number of Distinct groups	15	24	50

Though XRF A was the least sensitive of the three configurations XRF configurations previously tested in this study the overall discrimination power for the set remained comparable, demonstrating the utility of XRF on tape examinations. It is worth pointing out that XRF could only classify the samples into 24 distinct groups while LA-ICP-MS distinguished 50 groups. However, the additional groups differentiated by LA-ICP-MS were sub-groups that shared very similar elemental profiles not distinguishable by SEM-EDS or XRF.

4.5. Conclusions- XRF analysis of electrical tapes

In this study, the discrimination power was evaluated for three XRF configurations for the characterization of electrical tape backings. Forty electrical tape backings were characterized on an X-ray beam coupled to an SEM-EDS (Instrument A) and on a stand-alone XRF instrument (Instrument B). The two systems had comparable discrimination potential to LA-ICP-MS (81.5% and 90.1% vs. 84.6%, respectively). An

additional standalone μ-XRF instrument (Instrument C) was used to evaluate the discrimination power between configurations. On a subset of 12 tapes, XRF C had a discrimination power of 90.9%, compared to 84.8% and 97.0% for XRF A and B, respectively. All three configurations had discrimination power greater than SEM-EDS for the subset. Inter-day, intra-day and intra-roll variation were lower than the inter-roll variations, demonstrating the utility of XRF for elemental comparisons of backings. With a set of eighty-nine samples, XRF achieved a discrimination power of 91.0%, demonstrating that with the full set of tapes the method provided lower discrimination than LA-ICP-MS (94.2%) and consistently superior to SEM-EDS (87.6%).

Utilizing XRF requires optimized sample preparation and precaution to avoid interference from the sample holders; the penetration depth of XRF can cause significant errors if the instrument measures artifacts underneath the backings. Integrating the peaks and calculating the signal-to-noise ratios for the tape samples allows for determination of the elements that are present in a sample above the limit of detection, which aids in selection of elements for use in semi-quantitative intensity ratio calculations. Performing ANOVA and Tukey-Kramer analysis on the signal intensities aids in determining which sample pairs are significantly different and provides a more objective comparison method than spectral overlay alone.

As a result, XRF is determined to be a viable technique for the forensic examination and comparison of electrical tape backings. The study suggests that the instrumental configuration of the XRF system plays a critical role on the overall performance of the method. For instance, the stand-alone XRF system evaluated in this study provided superior capabilities in comparison to the X-ray tube attached to the SEM system. The main factors influencing the analytical performance of the XRF systems are the detector settings and the size of the beam. Since the sample size is not as limited in tape examinations as in other trace evidence materials, the examiner can afford to use collimators with large spot size to improve detection limits and to reduce acquisition times. The method offered added value to SEM-EDS elemental analysis, regarding the compositional information, classification capabilities and discrimination. Although XRF is a less sensitive and selective technique than LA-ICP-MS, among the samples tested in this study, the two techniques provided comparable discrimination and informing power. Therefore, XRF provides an efficient, fast and cost-beneficial method for analyzing and comparing electrical tape samples.

5. Overall Conclusions and Future Work

The long-term goal of this research is to strengthen the means by which tape evidence is analyzed and interpreted in forensic laboratories, particularly concerning the development of objective criteria to inform and support the examiners' opinions. In this project, we aimed to improve two particular areas of interest to tape examiners; physical fits of duct tapes and elemental analysis of electrical tapes. Our project was focused on two main hypotheses. First, the development of a systematic quantifiable metric on duct tape ends can offer measurable means to assess the quality of a fracture fit and assess the weight of that evidence. Second, that X-ray fluorescence can provide superior information and discrimination of electrical tape backings than elemental methods currently in use in crime labs (SEM-EDS). This research provides evidence that the methods developed in the course of this study are effective in demonstrating the value of physical fits in duct tape and that XRF provides several advantages for the analysis of electrical tape backings.

During the research, we accomplished a primary goal to establish a systematic and quantitative approach to duct tape physical fits and provided empirical data to support the opinions of examiners through a variety of statistical assessments that can be used to illustrate the quality of the fit.

A comparison method was developed for the examination of duct tape ends that provides, for the first time, a systematic approach to quantify the quality of a tape physical fit. Advantages of this approach are:

a) Easy to implement – no additional sample preparation is required, and the protocol is straightforward and uses microscopes widely available in forensic laboratories.

b) Structured and reproducible – the use of consistent features across the tape fracture and defined smallest comparison sub-unit ensures the examiners will follow a systematic method in which they will be observing at the same comparison areas during an examination. The study has demonstrated consistency among examiners better than 10%. However, a more extensive inter-examiner study is needed to demonstrate reproducibility fully.

c) Provides means for transparent peer review process– the documentation of specific areas of alignment, or the lack of, and the respective scores facilitates a transparent procedure during a discussion of results among examiners. The method also provides means to

present the evidence in court and explain the rationale of their conclusions with demonstrable criteria.

d) Provides a quantitative metric, defined as the ESS, that serves to inform the examiner's opinion–the study of the distribution of ESS metrics on populations of known origin permits an assessment of how much weight a physical fit holds relative to the quantitative score.

e) Provides probabilistic assessments that support the examiners conclusion-the use of SLR is one of several probabilistic approaches that can be applied with the documentation of an EES.

f) Permits a better understanding of factors that can affect the accuracy of physical fits-the use of quantifiable metrics offered means to evaluate if factors such as quality of the tape, separation method and stretching level affect the occurrence of errors.

A blind validation set of 2280 tape samples of "known origin" was examined to estimate performance measures. Since the effect of some factors was unknown at the beginning of the research, we took a conservative approach to modify one independent variable at a time to explore the effect of quality of tape, separation method and stretching on the accuracy. The results indicate that separation method nor moderate stretching do not contribute substantial differences in ESS distributions and accuracy. Extensive degrees of stretching did result in more misclassifications of samples, particularly increasing the number of false negatives. Tape grade did contribute the most in the tearing patterns, features observed across the fractured edges, and the distribution of ESS across true positive and true negative populations. The true matching tapes in the mid-grade sets generally produced scores between 90-100%, while the true matches in the low and high-quality sets received scores between 70-90%. The shift in score distributions of true matches did not correlate to a decrease in accuracy for the low and mid-quality sets (accuracy higher than 99%). However, for the high-quality tape set the incidence of false negatives lower accuracy at 84.9%. A key result from this study was that no false positives were observed for any set – demonstrating that even in the event of distortion examiners are unlikely to report a fit between edges that did not separate from the same source. With this preliminary knowledge base, future studies will include a more extensive assessment of the significance of those factors in the accuracy of physical fit examinations.

Similarity metrics were useful to use exploratory methods of data analysis (boxplots, histograms). ROC curves, SLR, and Tippet plots were useful to estimate rates of misleading evidence, evaluate the discrimination power of the method, and assess the value of the score-based likelihood ratios with regard to each tape grade. These statistical tools consistently indicate that there is a high accuracy rate in physical fit examinations, and on average, ESS values higher than 80 provide estimates of strong certainty of a match, while ESS values lower than 20 provide strong support for the conclusion of a non-match. The data also indicates that when values of ESS are between 40-60 an inconclusive result is likely, as the ESS values in that range did not contribute strong support towards either a match or non-match conclusion. For example, based on the experimental data, an examiner who observes an ESS of 90 for a tape edge has stronger support in the conclusion of a match compared to a tape edge that resulted in an ESS of 70. Likewise, more weight to a non-match conclusion can be offered for a score of 10 than for a score of 40. The use of the SLR provides further probabilistic statements. This quantifiable approach offers more objective criteria that support the examiner opinion that the observed features are substantial enough to declare a match or a non-match.

Another primary goal of this study was to develop and validate an analytical protocol for XRF in the analysis of electrical tapes. This research sets an important foundation that provides evidence that XRF is a superior analytical tool than SEM-EDS for the characterization and comparison of electrical tape backings. These results are relevant for the forensic community because XRF is a common tool available in crime laboratories, but currently underutilized for tape examination. The use of new sensitive elemental analysis tools enhances not only the discrimination capabilities and certainty of comparative examinations, but also increases the capacity for classification of samples for investigative leads where no comparison rolls are yet available. In this regard, XRF was able to increase classification for a validation set of 90 tapes from 15 groups to 24 groups. The discrimination power was also increased from 87.6% to 91.0%, which is comparable to ICP-MS capabilities.

Advantages of XRF over LA-ICP-MS include lower cost of acquisition and more widespread availability in forensic laboratories. Additionally, XRF advantages include:

a) Reliability of the method and the results – elemental profiles and classifications reported for all three XRF configurations were corroborated by multiple examiners in studies conducted over the course of several years utilizing different analytical techniques.

b) Informative – XRF was able to detect 14 major elements in tape commonly used to distinguish between samples originating from different sources. Despite not detecting as many elements as LA-ICP-MS (29 elements), XRF provided detection of additional elements as compared to SEM-EDS (8 elements), and the X-ray data could be used to classify samples.

c) Reproducible – no false positives were reported on duplicate controls samples analyzed within the same day and during a month (intra-day and inter-day variations). Also, intra-roll variation was found to be significantly smaller than the variation between different samples.

d) Increased discrimination power – in all three configurations, XRF demonstrated superior discrimination capacity to SEM-EDS, and in one configuration outclassed LA-ICP-MS on the samples analyzed.

e) Minimal sample preparation – adhesive removal is necessary for XRF analysis but is a fast and simple process.

f) Minimally destructive – the only alteration to the sample is the removal of the adhesive, and so samples can be utilized for further analysis.

g) Useful for typical sample sizes received in casework – the measurement area for XRF can range from 25 µm to nearly 1 cm diameter; however, tape samples received in forensic cases usually are much larger than typical trace evidence samples. Therefore, the method fits for purpose.

h) Easy to operate – the software used for XRF analysis is user-friendly with automated algorithms for peak identification and integration as built-in features.

i) Fast – analysis of a tape sample for XRF can take between 20-200 seconds per replicate, and the instruments can often be programmed for automatic sequences that can be run unattended.

j) Easy to implement – XRF instruments are already prevalent in many trace evidence laboratories for analysis of materials like glass, metals, and soil. With minimal required sample preparation, relatively large sample sizes, and high discrimination capacity from

the elemental characterization, XRF is anticipated to be adopted in tape examinations in the near future,

Overall, the study resulted in the development and validation of two robust methodologies for the analysis of tape evidence. The developed protocols and statistical assessments will support the opinions of examiners analyzing duct tape and electrical tape. Nonetheless, as with any other new methodologies in forensic science, the methods still need to undergo a more thorough validation process, including the expansion of datasets and a series of interlaboratory tests among examiners.

Ongoing research in our group includes further collaboration with forensic practitioners to cross-validate XRF methodology under more diverse instrumental configurations and conduct blind interlaboratory studies to generate a basis for the development of a standard test method. Our group is also collaborating with statisticians and computer scientists to develop automatic algorithms and alternative statistical models for the assessment of physical fits of trace evidence materials. Digital photographs of the tape edges have been collected for these purposes, and scanned images are being taken.

6. References

1. Scientific Working Group on Materials Analysis (SWGMAT). A 2012 Survey Regarding the Status of Forensic Tape Analysis. http://www.asteetrace.org/WebPages.cfm?action=viewCat&BIZ_UNL_id=41582.

2. Scientific Working Group on Materials Analysis (SWGMAT). Guideline for forensic examination of pressure sensitive tapes. *J American Soc Trace Evidence Ex.* 2011;2(1):88-97

3. Scientific Working Group on Materials Analysis (SWGMAT). Guideline for assessing physical characteristics in forensic tape examinations. *J American Soc Trace Evidence Ex.* 2014;5(1):34-50.

4. Mehltretter AH, Bradley, M. J. Forensic Analysis and Discrimination of Duct Tapes. *J American Soc Trace Evidence Ex*; 2012; 3(1): 2-20

5. LaPorte K, Weimer R. Evaluation of Duct Tape Physical Characteristics: Part I – Within-Roll Variability. *J American Soc Trace Evidence Ex*. 2017; 7(1): 15-34

6. Mehltretter AH, Wright DM, Smith MA. Variation in duct tape products over time: Physical measurements and adhesive compositional analysis by Fourier transform infrared spectroscopy. *Forensic Chemistry*. 2017; 4:1-8.

7. Keto R.O. Forensic Characterization of Black Polyvinyl Chloride Electrical Tape. *Crime Lab Digest.* 1984;11(4):71-74.

8. Okochi N., Shirai M., Nishijima K. Pressure-sensitive adhesive tape and pressure-sensitive adhesive composition. U.S. 20060182958A, in Nitto Denko Corporation, 2006.

9. Nakagawa Y., Ishiguro S., Natsume M., Takada, S., Ooyama, K., Yanagida, H. Adhesive tape and substrate for adhesive tape. U.S. 6773806B, in Nitto Denko Corporation, 2004.

10. Tucker P., Pyun E. Electrically insulating film backing. U.S. 5498476, in Minnesota Mining and Manufacturing Company, 1996.

11. Fujiwara S., Shirai M., Kawaguchi Y., Nakagawa Y., Ukei, H. Pressure-sensitive adhesive tape or sheet. U.S 20040126542A, in Nitto Denko Corporation, 2004.

12. Mehltretter A.H., Bradley M.J., Wright D.M. Analysis and Discrimination of Electrical Tapes: Part II. Backings. *J For Sci*. 2011;56:1493-1504.

13. Goodpaster JV, Sturdevant AB, Andrews KL, Brun-Conti L. Identification and comparison of electrical tapes using instrumental and statistical techniques: I. Microscopic surface texture and elemental composition. *J For Sci.* 2007;52:610–29

14. Martinez-Lopez C, Trejos T, Mehltretter AH, Almirall JR. Elemental analysis and characterization of electrical tape backings by LA-ICP-MS. *For Chem* 2017;4:97-107.

15. Martinez-Lopez C, Sakayanagi M, Almirall J. Elemental analysis of packing tapes by LA-ICP-MS and LIBS. *For Chem*. 2018;8:40-48.

16. Prusinowski M, Mehltretter A, Martinez-Lopez C, Almirall J, Trejos T. Assessment of the utility of X-ray Fluorescence for the chemical characterization and comparison of black electrical tape backings. *For Chem*. 2019;13:100146.

17. Martinez-Lopez C, Trejos T, Coulson S, Goodpaster J, Igowsky K, Kuczelinis F, Mehltretter A, Pollock E, Simmross U, Weimer R, Weis R, Almirall J. Interlaboratory evaluations of the performance of elemental analytical methods for the forensic analysis and comparisons of electrical tapes. *For Chem.* 2019;12:66-77.

18. Kuckzelinis F, Weis P, Bings NH. Forensic comparison of PVC tape backings using time resolved LA-ICP-MS analysis. *For Chem*. 2019;12:33-41.

19. Nelson D.F. Illustrating the Fit of Glass Fragments. *J of Criminal Law, Criminology, and Police Science*. 1959;50(3):312-314.

20. Baca A.C., Thornton J.I., Tulleners F.A. Determination of Fracture Patterns in Glass and Glassy Polymers. *J For Sci.* 2016;61(S1):S92-S101.

21. Quinn GD. Fractography of ceramics and glasses. *NIST Special Publication*. 2016:960-26e2.

22. Funk HJ. Comparison of Paper Matches. *J For Sci.* 1967;13(1):137 – 143.

23. Von Bremen UG. Laser Excited Luminescence of Inclusions and Fibers in Paper Matches. *J For Sci* 1986;31(2):455-463.

24. T. Tsach, S. Wiesner, Y. Shor. Empirical proof of physical match: Systematic research with tensile machine. *For Sci Int.* 2007;166:77-83.

25. Walsh K, Gordon A. Pattern matching of a paint flake to its source. *AFTE Journal.* 2001;33(2):143-5.

26. Vanderkolk JR. Identifying Consecutively Made Garbage Bags Through Manufactured Characteristics. *J For Ident.* 1995;45(1):38 – 50.

27. Von Bremen UG and Blunt LKR. Physical Comparison of Plastic Garbage Bags and Sandwich Bags. *J For Sci.* 1983;28(3):644-654.

28. Pierce D.S. Identifiable Markings on Plastics. *J For Ident*. 1990;40(2):51-59.

29. Bisbing R et. al. A Fingernail Identification. *AFTE Journal* 1980;12(1):27- 28.

30. White R, Arrowood M. Ultraviolet Fluorescence and a Physical Match. *AFTE Journal*. 1975;7(2):105-106.

31. Laux D.L. Identification of a Rope by Means of a Physical Match Between the Cut Ends. *J For Sci* 1984; 29(4):1246-1248.

32. Perper J, Prichard W and McCommons P. Matching the Lost Skin of Homicide Suspect. *For Sci International.* 1985;29:77-82.

33. Miller J, Kong H. Metal fractures: matching and non-matching patterns. *AFTE Journal*. 2006;38(2):133-165.

34. Christensen AM, Sylvester AD. Physical Matches of Bone, Shell and Tooth Fragments: A Validation Study. *For Sci International.* 2008;53(3):694-698.

35. Bradley MJ, Gauntt JM, Mehltretter AM, Lowe PC, Wright DM. A Validation Study for Vinyl Electrical Tape End Matches. *J For Sci.* 2011;56(3):606-611.

36. Bradley MJ, Keagy RL, Lowe PC, Rickenbach MP, Wright DM, LeBeau MA. A Validation Study for Duct Tape End Matches. *J For Sci.* 2006;51:504-508.

37. McCabe KR, Tulleners FA, Braun JV, Currie G, Gorecho EN. A Quantitative Analysis of Torn and Cut Duct Tape Physical End Matching. *J For Sci.* 2013; 58:S34-S42.

38. Tulleners F, Braun J. The statistical evaluation of torn and cut duct tape physical end matching. National Institute of Justice. 2011. Jul. Report No. 235287.

39. Agron N, Schecter B. Physical Comparisons and Some Characteristics of Electrical Tape. *AFTE Journal*. 1986;18(3):53 – 59.

40. Weimar B. Physical Match Examination of Adhesive PVC-Tapes: Improvement of the Conclusiveness by Heat Treatment. *AFTE Journal.* 2008;40(3):300-302.

41. Weimar B. Physical match examination of the joint faces of adhesive PVC-tapes. *AFTE Journal.* 2010;42(3):271-7.

42. National Academy of Science (NAS). National Academy of Sciences, Strengthening Forensic Science in the United States: A Path Forward. 2009.

43. President's Council of Advisor on Science and Technology (PCAST). Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/otsp/PCAST/pcast_forensic_science_report_final.pdf

44. OSAC Research Needs Trace Materials Subcommittee. Development of Quantitative Assessment and Evaluation of Error Rates in Physical Fit Determinations of Trace Materials. 2018. https://www.nist.gov/sites/default/files/documents/2019/02/01/osac_research_needs_aassessment_form_materials_sc_validation_of_physical_matches.pdf

45. Yekutieli Y, Shor Y, Wiesner S, Tsach T. Physical matching verification; Final Report to United States Department of Justice on Grant 2005-IJ-R-051. National Criminal Justice Reference Service. 2012.

46. Ristenpart W, Tulleners F, Alfter A. Quantitative algorithm for the digital comparison of torn and cut duct tape; Final Report to the National Institute of Justice Grant 2013-R2-CX-K009. University of California at Davis. 2017.

47. Jayaprkash P. Practical relevance of pattern uniqueness in forensic science. *For Sci International.* 2013;231:403.e1-403.e16.

48. Fawcett T. An introduction to ROC analysis. *Elsevier.* 2006; 861-874.

49. *Frigge M, Hoaglin, DC, Iglewicz B. Some Implementations of the Boxplot. The American Statistician. 1989;43(1):50–54.*

50. Venables WN, Ripley BD. Density Estimation. *Modern Applied Statistics with S.* Springer, 4th edition. 2002.

51. Zucchini, Walter. Applied smoothing techniques Part 1: Kernel Density Estimation (PDF). 2003. Retrieved at http://staff.ustc.edu.cn/~zwp/teach/Math-Stat/kernel.pdf.

52. Bunch S, Wevers G. Application of likelihood ratios for firearm and toolmark analysis. *Science & Justice.* 2013;53:223-229.

53. Leegwater AJ, Meuwly D, Sjerps M, Vergeer P, Alberink I. Performance study of a score-based likelihood ratio system for forensic fingermark comparison. *J For Sci.* 2017;62(3):626-640.

54. Hepler A, Saunders C, Davis L, Buscaglia J. Score-based likelihood ratios for handwriting evidence. *For Sci International.* 2012;219:129-140.

55. Gonzalez-Rodriguez J, Fierrez-Aguilar J, Ramos-Castro D, Ortega-Garcia J. Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. *For Sci International*. 2005;155:126-140.

56. Zadora G, Ramos D. Evaluation of glass samples for forensic purposes – an application of likelihood ratios and an informational-theoretical approach. *Chemometrics and Intelligent Laboratory Systems*. 2010; 102: 63-83.

57. Zadura, G., Martyna, A., Ramos, D., Aitken, C. Statistical analysis in forensic science: evidential value of multivariate physiochemical data. Wiley-Blackwell, Chichester. 2014.

58. Peter Gill, James Curran, Cedric Neumann. Interpretation of complex DNA profiles using Tippett plots. *For Sci International: Genetics Supplement Series.* 2008;646–648.

59. ASTM E3085-17, Standard Guide for Fourier Transform Infrared Spectroscopy in Forensic Tape Examinations, ASTM International, West Conshohocken, PA, 2017, www.astm.org

60. Skoog, D. A.; Crouch, S. R.; Holler, F. J. Principles of instrumental analysis; Thomson Brooks/Cole: Belmont, CA, 2007.

61. Mehltretter, A. H., Bradley, M. J. and Wright, D. M. Analysis and Discrimination of Electrical Tapes: Part I. Adhesives. *J For Sci*. 2011; 56: 82–94.

62. Golstein J, Newbury D, Michael J, Ritchie N, Scott JH, Joy D. Scanning Electron Microscopy and X-Ray Microanalysis. Springer, NY. 2018.

63. Brouwer, P. Theory of XRF. *PANalytical* B.V., 2003.

64. Margui, E., Greiken, R. *X-Ray Fluorescence Spectrometry and Related Tecniques: An Introduction*. Momentum Press. 2013.

65. Drake, L. XRF User Guide, http://www.xrf.guru/styled-12/page40/index.html.

66. Ernst T., Berman T., Buscaglia J., Eckert-Lumsdon T., Hanlon C., Olsson K., Palenik C., Ryland S., Trejos T., Valdez M., Almirall J.R. Signal-to-noise ratios in forensic glass analysis by micro X-Ray fluorescence spectrometry. *X-Ray Spectrometry*. 2014;43:13-21.

67. ASTM E2926-17. Standard Test Method for Forensic Comparison of Glass Using Micro X-ray Fluorescence (μ-XRF) Spectrometry. ASTM International, West Conshohocken, PA. 2017. www.astm.org

68. Gelman A. Analysis of Variance. 2006. Accessed at http://www.stat.columbia.edu/~gelman/research/published/econanova3.pdf.

69. Gerhart FJ and Ward DC. Paper Match Comparisons by Submersion. *J For Sci* 1986; 31(4):1450-1454.

70. Ausdemore MA, Neumann C, Saunders CP, Armstrong DE. Two-stage approach for the inference of the source of high dimension and complex chemical data in forensic science. 2018. Submitted to *Journal of American Statistical Association*. arXiv:1804.01049.

71. Hendricks JH, Neumann C, Saunders CP. Quantifying the weight of fingerprint evidence using an ROC-based approximate Bayesian computation algorithm. Submitted to *Journal of the Royal Statistical Society C*. 2018. arXiv:1803.10121.

72. Armstrong DE. Development and Properties of Kernel-based Methods for the Interpretation and Presentation of Forensic evidence. 2017. Dissertation, South Dakota State University. https://openprairie.sdstate.edu/etd/2175/.

73. Trejos T., Corzo R., Subedi K., Almirall J. Characterization of toners and inkjets by laser ablation spectrochemical methods and Scanning Electron Microscopy-Energy Dispersive X-Ray Spectroscopy. Spectrochimica Acta Part B. 2014;92:9-22.

Appendix 1.
Publication of Results for the Elemental Analysis of Electrical Tapes
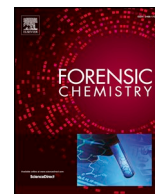
Assessment of the utility of X-ray Fluorescence for the chemical characterization and comparison of black electrical tape backings

Meghan Prusinowski[a], Andria Mehltretter[b], Claudia Martinez-Lopez[c], Jose Almirall[c], Tatiana Trejos[a],*

[a] Department of Forensic and Investigative Science, West Virginia University, Morgantown, WV 26506, United States
[b] Laboratory Division, Federal Bureau of Investigation, 2501 Investigation Parkway, Quantico, VA 22135, United States
[c] Department of Chemistry and Biochemistry and International Forensic Research Institute (IFRI), Florida International University, Miami, FL 33199, United States

HIGHLIGHTS

- XRF is a suitable tool for the forensic examination of electrical tape backings.
- XRF offers speed of analysis, minimal sample destruction, and high informing power.
- XRF elemental profile of electrical tape backings includes up to 14 elements.
- Intra-roll variability lower than variability between tapes of different origin.

ABSTRACT

Elemental analysis of electrical tapes is typically conducted by Scanning Electron Microscopy-Energy Dispersive Spectroscopy (SEM-EDS), although Laser Ablation-Inductively Coupled Plasma – Mass Spectrometry (LA-ICP-MS) recently showed to enhance the sensitivity of the determinations. In this study, the utility of X-ray Fluorescence (XRF) is compared to previously published SEM-EDS and LA-ICP-MS data. Three XRF systems were used to evaluate a range of configurations commonly available at crime laboratories. A set of 40 electrical tape backings known to originate from difference sources was used to assess the inter-roll variability, discrimination and classification capabilities of the method. The discrimination for this tape set increased from 78.8% achieved by SEM-EDS to 81.5–91.0% by XRF, depending on the instrumental configuration. In comparison, LA-ICP-MS achieved 84.6% discrimination on these tapes. The overall characterization, classification, and discrimination capabilities for this set improved as follows: SEM-EDS < iXRF < small spot size benchtop XRF with SDD, LA-ICP-MS < large spot size XRF. A set of 20 pieces of tapes collected from the same roll were analyzed for intra-roll variability. Duplicate control same-source samples were used to evaluate inter-day and intra-day instrument variability. No false exclusions were observed in the data set, demonstrating the within-sample variability and instrumental variability are relatively lower than the inter-sample variability. One concern of the method is the penetration depth of the X-ray beam beyond the target backing, requiring careful sample preparation to avoid interference from the adhesive or sample holder. XRF is a viable analytical tool for the forensic examination of electrical tapes, with advantages of speed of analysis, minimal destruction of the tape, and high informing power.

1. Introduction

Electrical tapes are subjected to forensic examinations in a variety of cases, including bombings and shootings. Forensic analysis of adhesive tapes aids in attribution or exclusion of potential sources. Because tapes are mass-produced, source associations cannot definitively establish that two items originated from the same source. However, the numerous physical and chemical differences among different formulations and manufacturers of tapes permit tapes to be grouped into distinctive classes, allowing for a high potential source exclusion rate.

The process for analyzing tape samples begins with a physical examination [1,2]. Electrical tapes consist of a backing layer and an

---

Table 1
Instrumental parameters for the XRF instruments used in this study.

	A (IXRF-SEM-EDS)	B (Thermo QuantX XRF)	C (Bruker XRF)
Instrument	IXRF beam + JEOL JSM-6490LV + INCA x-sight EDS detector	Thermo ARL QUANT'X EDXRF	Bruker M4 Tornado XRF
X-ray source	Rh	Rh	Rh
Detector	SiLi (NCD)	SiLi (PCD)	Silicon Drift (SDD)
Spot size diameter	$\sim$150 µm	$\sim$1 cm	$\sim$25 µm
Voltage (kV)	50 kV	Low 12 kV, Mid 28 kV,High 50 kV	50 kV
Power (µA)	1000 µA	Low 600 µA, Mid 900 µA, High 1120 µA	500 µA
Working distance (mm)	17	54.1	10.5
Dead time (%)	12–18%	18–40%	$\sim$30%
Collection time (live seconds)	200	20	200

adhesive layer. Physical characteristics such as the width of the tape, color of the backing and the adhesive, texture and thickness of the backing are features that examiners utilize to compare tape samples. However, unlike duct tapes, electrical tapes are more limited in the types of patterns and colors commonly used by consumers. Therefore, the chemical composition of the layers can be evaluated by techniques such as Fourier Transform Infrared Spectroscopy, Pyrolysis-Gas Chromatography-Mass Spectrometry, and Scanning Electron Microscopy-Energy Dispersive Spectroscopy to enhance the informing power of the comparison tests [1].

Electrical tape backings are primarily polyvinyl chloride (PVC), and therefore chlorine is often found at high concentrations. The chemical bonds make PVC rigid under standard conditions, so plasticizers are added to increase the flexibility of the backing. Additional components such as stabilizers, fillers, and flame retardants modify the utility and quality of the tape. Some of the elements incorporated in this process originate from the pigments and fillers such as TiO_2, $CaCO_3$, and $BaSO_4$, and stabilizers such as lead, barium, and calcium salts [3–7]. Flame retardants can include antimony oxide, aluminum hydroxide, magnesium carbonate, as well as other halogens, phosphorous, and molybdenum compounds [5–7]. The relative compositions and concentrations of elements are expected to vary between products and manufacturers, and aid in distinguishing tapes from different sources. Prior studies have shown that elemental analysis is highly discriminating for differentiating among electrical tape backings [8], due to the variety of inorganic components added in the manufacturing process.

This study aims to compare the performance of three XRF instrumental configurations to SEM-EDS and LA-ICP-MS for the elemental analysis of electrical tape backings for forensic comparisons. XRF is an appealing analytical tool for the forensic analysis of tapes due to its almost non-destructive nature, fast analysis, ease of operation and relatively low acquisition cost. Both SEM-EDS and XRF detect X-rays with energies characteristic of the atomic structure of each element. However, XRF uses photons instead of an electron beam to generate the signal. SEM-EDS allows the examiner to characterize morphology of the surface and the particles that contribute to the elemental signal. Additionally, forensic questioned samples may have debris that can contribute to the elemental compositions and should be considered when conducting comparisons. XRF differs from SEM in that XRF is a more sensitive technique and can detect higher atomic number elements. SEM-EDS has a detection limit of around 0.1% by weight (or approximately 1000 µg/g), while XRF has a detection limit of about 10–50 µg/g [9]. However, because XRF uses higher energy photons, XRF penetrates deeper into the sample. Therefore, the different layers of backing and adhesive must be separated before analysis and mounted on surfaces that avoid interferences. XRF instruments can also utilize filters to improve sensitivity. Filters reduce the signal from the background and some interfering lines to improve the signal-to-noise ratio of the spectra [10].

LA-ICP-MS is a technique that utilizes a laser to remove a small amount (e.g., a few micrograms) of mass from a sample of interest [11].

This method offers high sensitivity and selectivity, with limits of detection in the low and sub-µg/g range for elements that aid in discriminating between tape samples [12,13]. However, LA-ICP-MS instruments are, in general, less available in forensic laboratories than SEM-EDS and XRF systems. As a result, this study aims to evaluate XRF as a viable method for elemental characterization of tapes as a cost-effective alternative for those forensic laboratories that routinely use the method for other materials such as glass and paint [9,11,14,15].

2. Methods and materials

2.1. Instrumentation and measurement parameters

The XRF instruments used in this study include a µ-X-beam (X-beam slide, IXRF, Austin TX) coupled to an SEM-EDS (JSM-6490LV, JEOL, Tokyo Japan), a standalone XRF (ARL QUANT'X EDXRF, ThermoFisher, Waltham MA) and a standalone µ-XRF (M4 Tornado, Bruker, Billerica MA) The systems are referred to as A (IXRF), B (Thermo QuantX), and C (Bruker M4 Tornado), respectively.

Table 1 lists the acquisition parameters for the instrument systems. The data from system B were acquired using three different filters between the X-ray tube and the sample to optimize the signal-to-noise ratio of the species of interest. A 0.13 mm aluminum filter was used for lower energy source X-rays (Low Z), while a 0.06 mm palladium filter and a 0.559 mm thick copper filter were used for the mid-energy (Mid Z) and high-energy region (High Z), respectively.

2.2. Sample collection and sample preparation

A subset of 40 electrical tape backings was selected from a set of 90 previously characterized by our research groups [8,12] and measured on two of the XRF systems (A and B). Sample ID numbers are as described by Mehltretter et al. [8] and Martinez et al. [12] (samples 1 to 25, 27 to 41). A subset of 12 out of the 40 tapes was also analyzed on system C. A minimum of three replicates were collected from each sample.

The tapes were cut into $\sim$1–2 cm^2 pieces, and the adhesive was removed from the backing using hexane. Three portions of each of the tape's backing were stored in pre-labeled plastic bags and coin envelopes. For samples measured using system A, the tape backings were coated with carbon ($\sim$100 Å) and then mounted on a transparent XRF film (Chemplex, NY) on top of an SEM sample holder. To prevent interferences from the sample holder, the XRF film was suspended on an open space of the holder of about 1 cm diameter by 2 cm depth. Samples measured on system B were placed directly on top of the holder aperture (approximately 2 cm diameter aperture), with a beryllium window on top of the tape to reduce signal contributed by X-ray interaction with the chamber material. For instrument C, the samples were placed atop a graphite planchet. Differences in sample mounting were a result of the specific structure of chambers, sample stages, and holders.

In addition, a roll of electrical tape (Super 33+, Scotch 3 M, Saint

Paul MN) was selected to evaluate the intra-roll variability on instrument A. The entire roll was unwound, and 20 pieces of ~1–2 cm were removed every ~38 in. The samples were prepared following the same procedure described above for instrument A.

Daily performance quality control tests used a copper/aluminum standard for calibration of the energy scale (x-axis). The calibration maintained the Cu Kα and Al Kα emission peaks within 5 eV of their literature values and monitored the peak resolution. Several tape samples were analyzed in duplicate as a daily monitor of instrumental variation over time. Duplicate controls, consisting of a piece of tape from the same roll, were analyzed as a daily monitor of instrument variation over time. Three replicate measurements were acquired for each duplicate control sample. The NIST standard reference material 1831, a soda-lime glass, was measured for instruments A and B every day as a QC control to monitor variability of the intensity of the iron and strontium peaks. Instrument C was monitored using manganese and zirconium for a daily control as directed by the manufacturer. Manganese is used as a standard element to determine the resolution of the detector. Zirconium is used during energy channel calibration of the detector because it falls approximately in the middle of the spectra.

2.3. Data analysis

The two approaches to comparing the data collected in this study were spectral overlay, and a semi-quantitative method in which the integrated areas under the peaks were calculated and compared statistically utilizing Analysis of Variance (ANOVA) and Tukey-Kramer post-hoc test. Spectral overlay is commonly used in forensic spectrochemical analysis to quickly distinguish between samples of differing composition. For XRF analysis, spectral overlay allows the examiner to pre-screen the spectra and compare the elemental profile of the samples. The sample origin was kept blind to the analyst. Differences between two samples are based on presence versus absence of characteristic peaks, or differences in the relative heights and shapes of the peaks. The signal-to-noise ratios are a criteria used in identification of peaks that would be relevant for comparative analysis as shown by Ernst et al. [15]. For a peak to be distinguishable, none of the replicate measurements should overlap any of the replicate measurements of the comparison sample. For example, this is the equivalent of a minimum-maximum range established for each relevant peak height. For the samples to be distinguishable, at least one element in the spectra needs to show no overlap in the peak range between the two samples. In this study, each pairwise comparison between tape samples was conducted with all the replicates to account for the variability within the sample [16].

A semi-quantitative approach can aid in distinguishing between samples, particularly those that were otherwise indistinguishable by spectral overlay. For this study, only data from samples collected on instrument A were used in the semi-quantitative assessment. The spectra were normalized prior to comparison to minimize the potential effects of tape thickness differences. The normalization involved dividing the spectra's counts at each energy interval by the sum of the overall counts of the individual spectra and multiplying by a factor of 100000. The integrated peak areas and the signal-to-noise ratios were calculated for each replicate spectra. The signal-to-noise ratios were used to determine which elements were above the limit of detection prior to statistical analysis [15]

ANOVA is a parametric statistical approach used to compare means between several independent groups. The one-way ANOVA test as used in this study looked at the between-samples variability and the within-sample variability for each particular element or elemental ratio. An ANOVA test will show if at least two samples are significantly different from each other, but will not specify which samples. A follow up test

Table 2
Energy ranges (keV) for select elements.

Elements	Prepeak	Peak	Postpeak
Al (Kα)	1.26–1.42	1.42–1.66	N.A
Si (Kα)	N.A	1.66–1.82	1.82–1.98
Cl (Kα)	2.31–2.53	2.53–2.75	N.A
Ca (Kα) + Sb (Lα)	3.24–3.49	3.49–3.74	N.A
Ba (Lα) + Ti (Kα)	4.31–4.41	4.41–4.62	4.62–4.72
Fe (Kα)	6.15–6.29	6.29–6.58	6.58–6.71
Zn (Kα)	8.36–8.50	8.50–8.78	8.78–8.92
Pb (Lα)	10.22–10.38	10.38–10.74	10.74–10.92
Br (Kα)	11.64–11.76	11.76–12.02	12.02–12.14
Cd (Kα)	22.32–22.68	22.68–23.42	23.42–23.78
Cr (Kα)	N.A	5.34–5.60	5.60–5.86
Mo (Kα)	17.16–17.29	17.29–17.56	17.56–17.69

must be run to determine which samples are different. For this purpose, the Tukey-Kramer test (also known as Tukey-Honestly Significant Difference) was used. This test performs each pairwise comparison of sample means, and considers all combinations of pairs to calculate the HSD statistic. The HSD statistic is compared to a critical value (i.e., at α = 0.05), to decide if the means are significantly different.

Spectral comparisons were conducted using the Iridium Ultra Software (IXRF, version 1.4, Austin TX), Microsoft Excel (Excel 2016, Redmond WA), open-source R (version 3.5.0), and JMP (SAS Institute, JMP version 14.0.1, Cary NC). Spectral overlay, integrated peak areas and signal ranges, calculated signal-to-noise ratios of elements, and ANOVA were used for comparison of spectra as described by Ernst et al. [15] and Martinez et al. [12]. The measured ranges in peak positions are reported in Table 2.

Several element peaks were not able to utilize pre- or post-peak ranges because the element peaks overlapped with other peaks in that energy range. Instead, a larger range was used on one of the side of the peak to calculate the background with the same number of channels as the analyte peak. Discrimination power (DP) was calculated as reported by Mehltretter et al. [8] and Martinez et al. [12].

3. Results and discussion

3.1. Evaluation of XRF for characterization and discrimination of tapes from different sources

3.1.1. Sample preparation considerations

Since the sensitivity of XRF is superior to SEM-EDS but lower than LA-ICP-MS methods, it was hypothesized that XRF would provide enhanced discrimination compared to SEM-EDS and a more cost-effective alternative to laser-ablation methods. With tape samples, the X-ray beam penetrates beyond the typical backing and adhesive thickness. The thickness of the electrical tape backings used in this study ranged from 96 to 198 μm, while XRF has the potential to penetrate up to several millimeters in depth depending on the atomic number of the element. As a result, samples should be prepared and exposed to the X-beam in such a way that prevents interference on the X-ray spectrum from any material underneath the layer of interest. Unlike SEM-EDS and LA-ICP-MS, XRF requires removal of the adhesive layer from the backing, and the samples mounted on a substrate that do not produce X-rays artifacts. For comparative analysis, it is therefore critical that the known and questioned samples are prepared in the same manner. Moreover, the direct comparison of XRF data from one laboratory to another represents a challenge for XRF due to differences in backgrounds, sensitivities, and detection capabilities. Future development of standard reference materials for tape analysis may mitigate some of the limiting factors between instrumental configurations. At this point,

Table 3
Samples grouped by SEM-EDS and XRF instruments A and B.

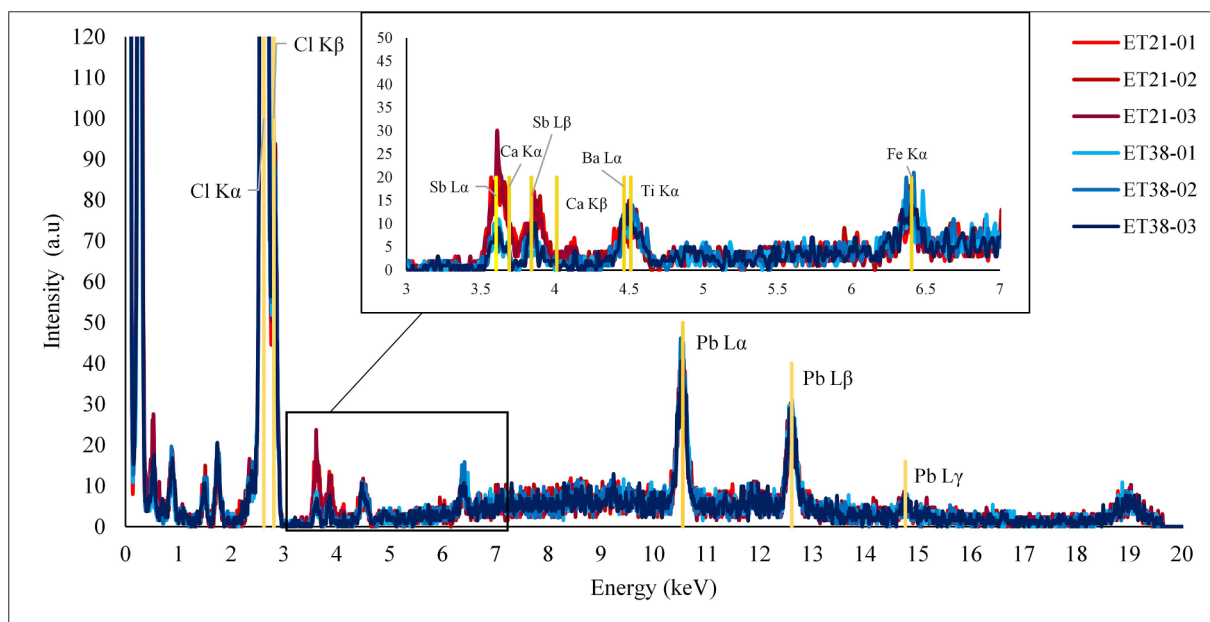| SEM-EDS Groups	Sample ID	XRF A Groups	Sample ID	Main differences by A	XRF B Groups	Sample ID	All Filters Main differences by B
1	4,8,32	1a	4		1i	4	Lower Ca than 1ii&1iii, higher Fe than 1ii
		1b	8,32	Higher Ca & higher Pb than 1a	1ii	8	
					1iii	32	Higher Ca than 1i&1ii
2	14,35,37	2	14,35,37		2	14,35,37	
3	21,38	3	21,38		3i	21	Higher Sb than 3ii
					3ii	38	Lower Ca, higher Fe than 3i
4	2,10,11,12,13,15,17,18, 19,20,23,24,25,27,28,39,41	4a	2		4i	2	Lower Zn than 4ii&4iii
		4b	10,11,12,13,15,17,18,19,20,23,24,25,27,28,39,41	Higher Zn than 4a	4ii	10, 17, 23, 24	Lower Fe, higher Pb, higher Sb than 4iii
					4iii	11, 12, 13, 15, 18, 20, 25, 27, 28, 39, 41	
					4iv	19	Lower Fe than 4i&4iii
5	16,29,30,34,36,40	5	16,29,30,34,36,40		5i	16,29,30,34,40	
					5ii	36	Contains Cr
6	1,5,7	6	1,5,7		6	1,5,7	
7	3,6,31	7a	3,6	Contains Ba/Ti	7i	3	
		7b	31	Higher Ca than 7a, no Ba/Ti	7ii	6	Higher Ca than 7i, lower Ba than 7i&7iii, contains Cd
					7iii	31	Higher Pb, higher Fe than 7i&7ii, no Ba
8	9, 33	8a	9		8i	9	
		8b	33	Lower Pb than 8a	8ii	33	Lower Pb, higher Fe, Zn than 8i
9	22	9	22		9	22	

Fig. 1. Spectra overlay comparison of tapes 21 (replicates 01–03) in red and 38 (replicates 01–03) in blue for instrument A.

forensic examiners should be aware of this limitation for sharing databases between laboratories.

3.1.1.1. Instrument A. For instrument A, the backings were suspended on X-ray film ~ 2 cm above the bottom of the sample holder. The diameter and depth of the holder were sufficiently large enough to prevent spectral interferences. Nonetheless, when the backing was sampled near the edges, artifact Cu peaks appeared in the spectrum from scattering of the holder walls. As a result, the areas of the spectra for Cu Kα and Cu Kβ (8.05 and 8.90 keV respectively) were omitted during the inter-sample and intra-sample comparisons.

3.1.1.2. Instrument B. The stand-alone XRF system B is configured to deliver the X-beam from underneath the sample holder and is equipped with a holder with a circular aperture of about 2 cm wide. The tape was directly attached to the aperture edges, and a beryllium disc was placed on top of the sample to reduce escape and scattering from the X-ray beam. Signal interference from the holder was negligible with this configuration.

3.1.1.3. Instrument C. Samples for system C were mounted on graphite planchets. Blank measurements of the planchets showed a peak at ~ 15.7 keV which was also observed in all sample spectra. Therefore, this peak was not considered in the comparisons.
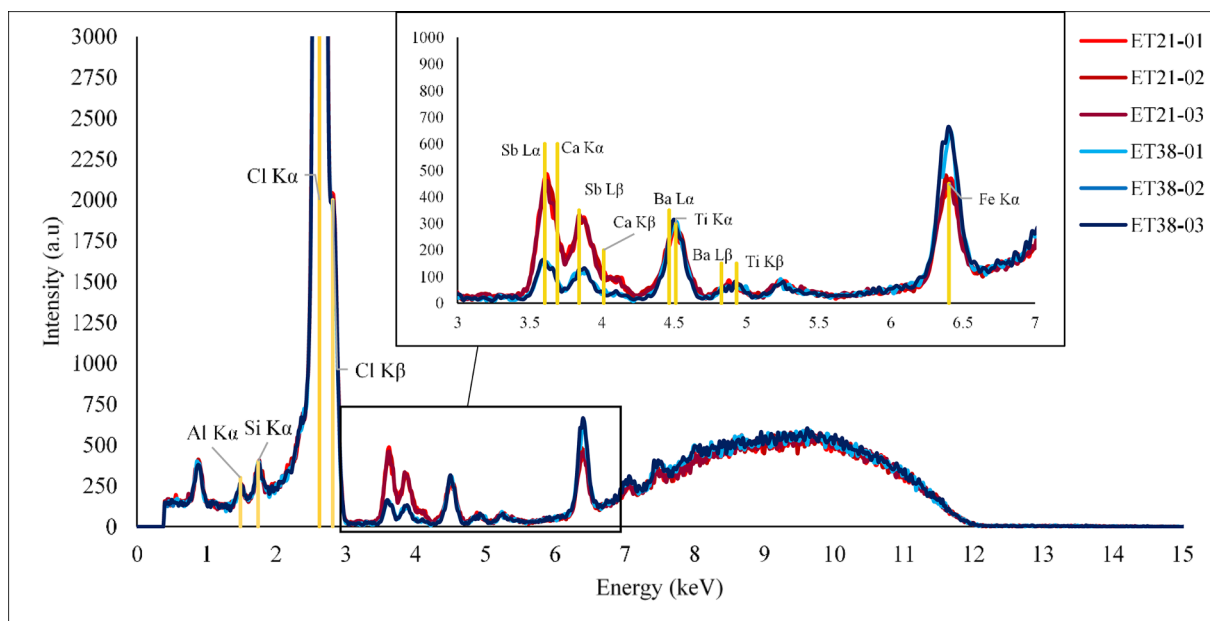

Fig. 2. Spectra overlay comparison of tapes 21 (replicates 01–03: red) and 38 (replicates 01–03: blue) for the low filter on instrument B.
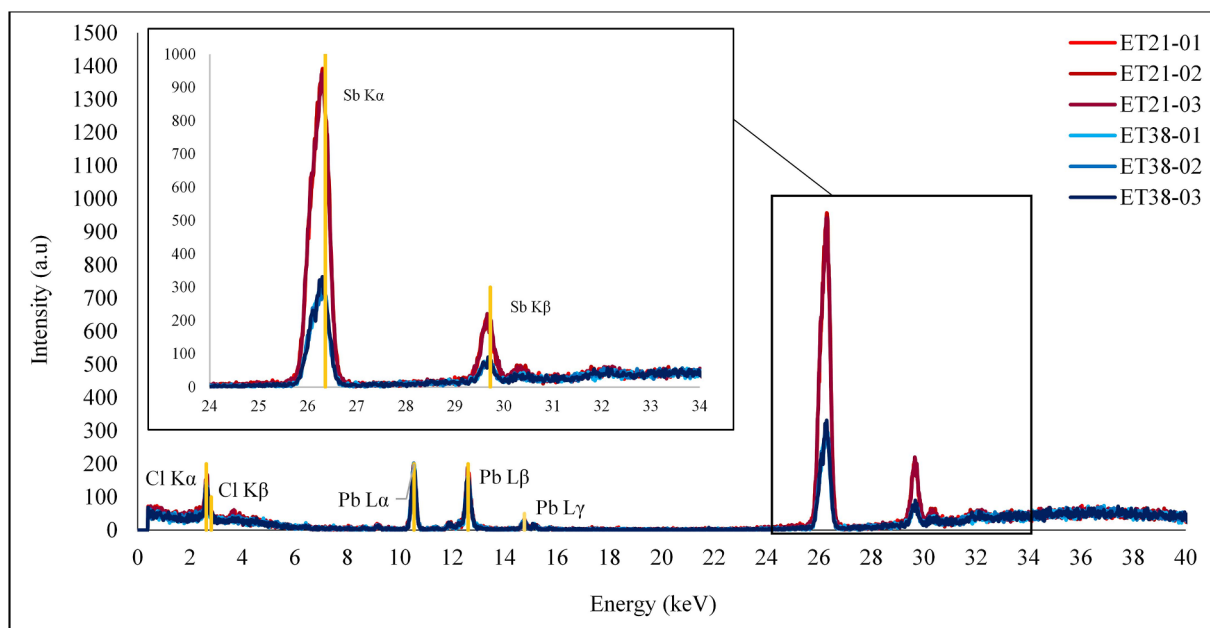
Fig. 3. Spectra overlay comparison of tapes 21 (replicates 01–03: red) and 38 (replicates 01–03: blue) for the high filter on instrument B.

3.1.2. Inter-source variability

A set of 40 tapes was selected from a previously characterized collection [8,12] purposely to include groups that show differences between various brands of tapes and groups that show similarity between same-brand tapes. The set consisted of tapes from 15 different brands and 30 different types of black electrical tapes. Table 3 summarizes the groups of indistinguishable samples as measured on XRF instruments A and B. Using spectral overlay, SEM-EDS alone could classify the 40 samples (780 comparison pairs) into nine distinctive groups, while instrument A further discriminated the samples into 13 groups, and instrument B was able to classify samples into 19 groups from data acquired using all three filters. The spectral overlay conclusions were verified by an additional analyst prior to the finalized groupings.

The superior performance of the stand-alone system B is the result of the detection configuration, the beam size, and the use of filters [10]. The larger the area and thickness of the Si(Li) crystal, the better the efficiency to capture photons. Also, unlike other trace evidence, the typical size of tape specimens allows analysis from relatively large areas, which can be advantageous on systems with adjustable spot size. The beam size used in B was 1 cm diameter, allowing for a sampling area at least 4400 times greater than the micro-beam used on the other XRF instruments. Another benefit from the increased sensitivity is that the acquisition time can be reduced to only 20 s per replicate, speeding up the overall analysis time.

3.1.2.1. Inter-source variability, instrument A.

Fig. 1 represents an example overlay of sample 21 shown in red and sample 38 shown in blue as measured by instrument A, previously not differentiated by SEM-EDS [8]. The elements found in the two samples include aluminum, chlorine, calcium, antimony, barium, titanium, iron and lead. Calcium, antimony, barium, and titanium were among the common elements found in the tapes measured and were oftentimes the most discriminating peaks between samples. However, standard resolution of EDS systems used here (0.1 keV), does not allow a baseline separation of the peak lines for antimony Lα (3.61 keV) and calcium Kα (3.69 keV) or the barium Lα (4.47 keV) and the titanium Kα (4.51 keV) peak lines. For samples 21 and 38, there was some degree of variation in intensity for the calcium and antimony peaks (∼3.5–4.2 keV), but it was not sufficient to distinguish between the tape samples (Fig. 1).

3.1.2.2. Inter-source variability, instrument B.

The filters placed between the X-ray tube and the specimen significantly reduced the background of the spectrum measured by instrument B. Filters allow the transmission of X-Rays that are below the absorption edge energy of the filter's element, while absorbing other source X-rays. The result is a low background above the filter's absorption edge energy that enhances the signal to noise ratio in that region. Three different filters were used on instrument B to optimize the spectral region of the elements of interest. The aluminum filter was used for the low atomic number elements (Low Z) to measure elements with Kα lines between 2 and 10

Table 4

Summary of the discrimination capability for the SEM-EDS, XRF and LA-ICP-MS.

	SEM EDS	XRF A	XRF B (low filter)	XRF B (mid filter)	XRF B (high filter)	XRF B (all filters)	LA-ICP-MS
Discrimination power (%)	78.8 (614 pairs out of 780)	81.5 (636 pairs out of 780)	90.0 (702 pairs out of 780)	81.8 (638 pairs out of 780)	86.1 (672 pairs out of 780)	90.1 (703 pairs out of 780)	84.6 (660 pairs out of 780)
Correct associations/Controls (%)	N.A	100 (328 pairs)	100 (4 pairs)	100 (4 pairs)	100 (4 pairs)	100 (4 pairs)	100 (279 pairs)
Number of elements detected	8	14	10	9	8	14	29
# Distinct groups	9	13	18	16	17	19	19

Table 5
Samples grouped for the three XRF instruments.

XRF A Groups	Sample ID	Detected elements	XRF B Groups	Sample ID	Detected elements	Main differences by B	XRF C Groups	Sample ID	Detected elements	Main differences by C
1	8,32	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb	1i	8	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb, Br	Contains Br	1I	8	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb, Br	Contains Br
			1ii	32	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb	Higher Ca/Sb than 1i	1II	32	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb	Higher Ca/Sb than 1I
2	21,38	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Pb	2i	21	Al, Si, Cl, Ca, Ba/Ti, Fe, Zn, Pb, Br	Higher Sb than 2ii	2I	21	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb, Br	Higher Ca/Sb than 2II
			2ii	38	Al, Si, Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb, Br	Higher Fe than 2i	2II	38	Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb, Br	
3	10,12,15,24	Cl, Ca/Sb, Zn, Pb, Mo	3i	10, 24	Cl, Ca/Sb, Fe, Zn, Pb, Mo	Lower Fe, higher Pb and Sb than 3ii	3	10,12,15,24	Cl, Ca/Sb, Zn, Pb, Mo	
			3ii	12, 15	Cl, Ca/Sb, Fe, Zn, Pb, Mo					
4	16,36	Cl, Ca/Sb, Zn, Pb, Mo	4i	16	Cl, Ca/Sb, Fe, Zn, Pb, Mo	Contains Cr	4I	16	Cl, Ca/Sb, Zn, Pb, Mo	Contains Cr
			4ii	36	Cl, Ca/Sb, Cr, Fe, Zn, Pb, Mo		4II	36	Cl, Ca/Sb, Cr, Zn, Pb, Mo	
5	3,6	Cl, Ca/Sb, Ba/Ti, Pb	5i	3	Cl, Ca, Ba/Ti, Fe, Zn, Pb	Higher Ca and lower Ba than 5i, contains Cd	5I	3	Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb	Higher Ca/Sb, higher Fe than 5I
			5ii	6	Cl, Ca, Ba/Ti, Fe, Zn, Pb, Cd		5II	6	Cl, Ca/Sb, Ba/Ti, Fe, Zn, Pb	

KeV (e.g., Ca, Fe, Zn). The mid-range filter (Mid Z) made of thick palladium was used with a tube operating voltage of 28 KV, and improved detection of Br, Sr, Cr, Pb, and Mo. The thick copper filter was used with a tube operating at a high voltage (50 KV) for optimal detection of Mo, Ba, Pb, and Sb (High Z). Using the filters resulted in superior sensitivity and selectivity for the tape samples. For tapes 21 and 38, the superior signal-to-noise ratio allows for the two tapes to be differentiated by the amounts of calcium/antimony and iron (Fig. 2).

In addition, the lack of resolution observed on EDS spectra for Ca and Sb peaks is clarified at the high Z, where Sb Kα and Kβ peaks, but not Ca, are present above 25 KeV. An example of this enhancement is shown in Fig. 3 for the Sb peaks corresponding to samples 21 and 38. Likewise, the difficulties in resolving Ba from Ti peaks in the low KeV region are no longer problematic in the high energy region. The added selectivity represents a valuable advantage because these elements are highly informative and discriminating in electrical tapes.

3.1.2.3. Comparison of instruments A and B for Inter-source variability. From the possible 780 possible comparison pairs from this set (n = 40) as seen in Table 4, the XRF method B discriminated 90.1% when using combined spectral overlay results from all three filters. The discrimination capabilities were higher than SEM-EDS (78.8%), XRF attached to SEM (method A) (81.5%) and LA-ICP-MS (84.6%) for the set [8,12].

The higher discrimination power of XRF instrument B compared to LA-ICP-MS was due to the capability of the instrument to classify the samples in group 4 (Table 3) into four separate classes based on the iron, lead and antimony content. Iron is more distinguishable using the aluminum filter, while antimony is resolved using the copper filter. Despite LA-ICP-MS having the capability to detect more elements, the method was only able to separate tapes 2, 27 and 28 from the rest of the samples in that group [12]. This is not surprising as iron detection is problematic in standard mass analyzers like the one used in this study [12]. Moreover, apart from tape 2, the tapes in that group originate from the same manufacturer, 3 M [8]. Due to similar manufacturing processes, the composition of the tapes in that subgroup were very similar by XRF and LA-ICP-MS. Both XRF methods and LA-ICP-MS distinguished sample 2, which was manufactured by Advance®, while SEM-EDS did not detect the different compositions [8]. It is worth noting that, except for group 4, XRF and LA-ICP-MS methods classified the tapes into the same classes for this set of tapes. Nonetheless, LA-ICP-MS still provided superior overall selectivity and sensitivity that allows identification and characterization of a larger number of elements.

3.1.2.4. Comparison of instruments A, B and C for Inter-source variability. A third XRF system was selected as an anticipated intermediate performance between A and B, and to represent another common XRF system found at crime laboratories. The second subset of 12 electrical tapes consisted of samples in 5 groups indistinguishable by SEM-EDS or XRF instrument A. Table 5 summarizes the spectral overlay group results as measured by the three XRF systems and the detected elements in each group.

Instruments B and C differentiated samples 21 and 38 by the calcium and antimony peaks as well as iron (Fig. 4). However, unlike system B, instrument C was unable to resolve between the peaks of calcium and antimony in the energy range utilized (Fig. 3).

XRF instruments A and C were unable to distinguish between samples originating from 3 M (samples 10, 12, 15, and 24). On the other hand, instrument B could distinguish between the two brands of tape represented in the four samples (3 M Super 33+ for samples 12 and 15, and 3 M Super 88 for samples 10 and 24) based on the iron and lead content. Instrument B at high Z was also able to distinguish the samples by the antimony content using the higher energy range that was not measured on the other instruments (Fig. 5).

Of the 66 possible comparison pairs for the twelve-tape subset, XRF method B had the highest discrimination power at 97% when using all
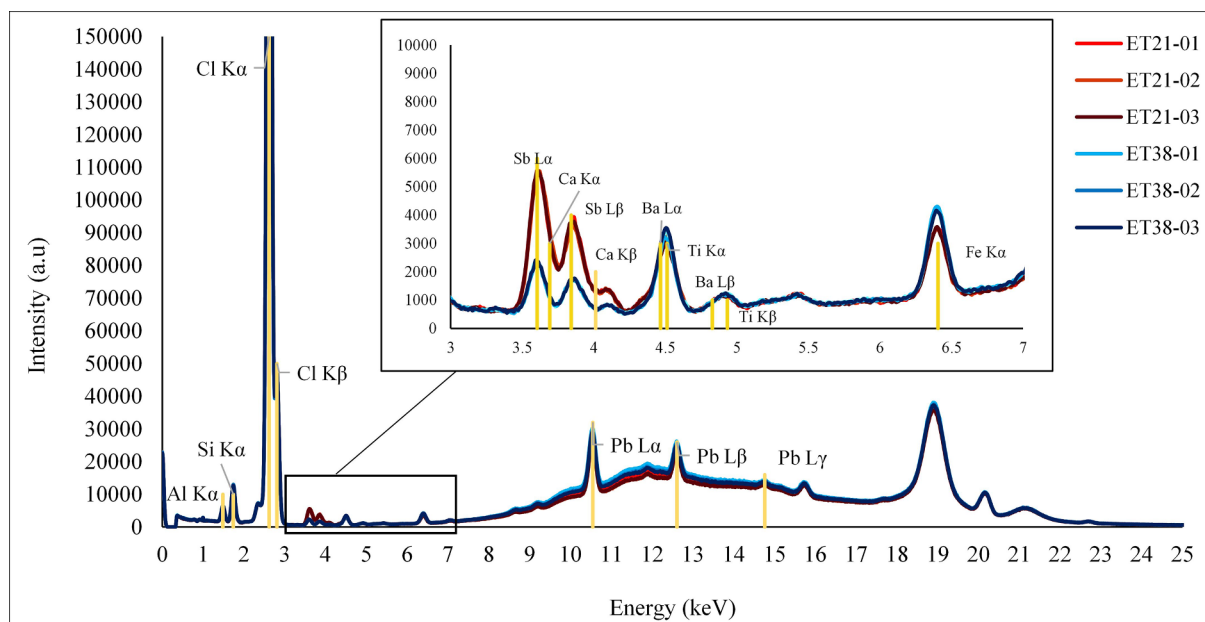
Fig. 4. Spectra overlay comparison of tapes 21 (replicates 01–03: red) and 38 (replicates 01–03: blue) for instrument C.

three filters. Overall, XRF had high discrimination potential even for samples that were very similar in composition (Table 6).

For the subset of 12 tapes, XRF system A performed comparably to SEM-EDS – the same tapes were indistinguishable for both instruments, resulting in a discrimination power of 84.8%. In contrast, system C and LA-ICP-MS separated these 12 tapes into the same groups, resulting in a discrimination power of 90.9%. While system B had the highest discrimination power of the three XRF instruments, instruments with this configuration are less commonly available in forensic laboratories as the other two.

3.2. Example of characterization and semi-quantitative evaluation of sample variability using instrument A

3.2.1. Comparison of intra-roll samples

ANOVA and the Tukey-Kramer test were used to determine which pairs of tapes had significant differences at each element. Quality control duplicates for sample 10 measured twice a day for each day of the study were analyzed to evaluate inter-day and intra-day variation for the instrument.

Sample ET10 was collected as the first and last sample (duplicate controls 10A-10N) for each day of the inter-roll study. The ANOVA and Tukey-Kramer results for the comparison of the Cl/Ca ratio, calculated using the integrated peak areas, are shown in Fig. 6. A graphical representation of the test can be depicted with circles, in which the center represents the mean of each group and the diameter is the HSD interval for that mean. If two circles overlap, then the two means are not different. The larger the overlap among circles, the more similar the groups are (larger p-values). If there is an angle of intersection less than 90 degrees between two circles, or no overlap in the circles, then the two samples are significantly different. The horizontal line represents the overall mean of all the samples, and the diamonds demonstrate the mean of each sample's replicate measurements. The measurements showed a low variation of the instrument over the course of the study. No significant differences in the means were found at $\alpha = 0.05$. The chlorine to calcium ratio was utilized to represent the worst-case scenario of within-sample variations, as it was observed by the spectral

overlay comparisons that the peak from the polyvinyl chloride showed relatively poor repeatability.

In addition to the duplicate control samples, a set of 20 electrical tape backings (labeled ETV01-ETV20) was evaluated for elemental variability of samples known to originate from the same roll. The data for intra-roll variation were collected over the course of two days with sample ETV01 measured three times on the first day and six times on the second day to constrain inter-day and intra-day variation of the instrument. The signal-to-noise ratios of Cl, Ca/Sb, Fe, Zn, Pb, and Mo were calculated using the peak ranges defined in Table 2. The elements that were found to have signal-to-noise ratios higher than 3 for the roll were Cl, Ca/Sb, Zn, and Mo. The three replicates within each sample were checked for outliers using Grubb's test. One duplicate sample (ETV02) was found to have a replicate that was beyond the typical variation seen in these samples. With only two replicates remaining, it could not be included in the statistical analysis, and so was removed. The remaining 21 samples were compared with spectral overlay and additionally by ANOVA followed by Tukey-Kramer analysis for the selected elements (Fig. 7).

For the intra-roll study, no significant differences in the Cl/Ca ratio or in the intensities of Zn and Mo were determined. All samples were correctly associated with the others from the same roll using the combination of spectral overlay and quantitative analysis. Intra-roll variations were found to be lower than inter-roll variations for electrical tape samples. Fig. 8 represents the Cl/Ca ratio and the signal intensities for elements Ca/Sb and Zn for samples classified into different groups on XRF A from the inter-roll study of the set of 12 similar tapes.

As illustrated in Fig. 8, samples 9 and 33 can be differentiated from all the other samples and each other utilizing the Cl/Ca ratio, which could not be accomplished with SEM-EDS [8]. Meanwhile, sample 31 can be differentiated by the amount of Ca and Sb, and sample 15 can be distinguished from most of the other samples by the Zn content. If one element is found to be significantly different between two samples, the tapes are found distinguishable. As such, relatively low within-sample and within-instrument variations are essential to recognize significant differences between samples from different sources.
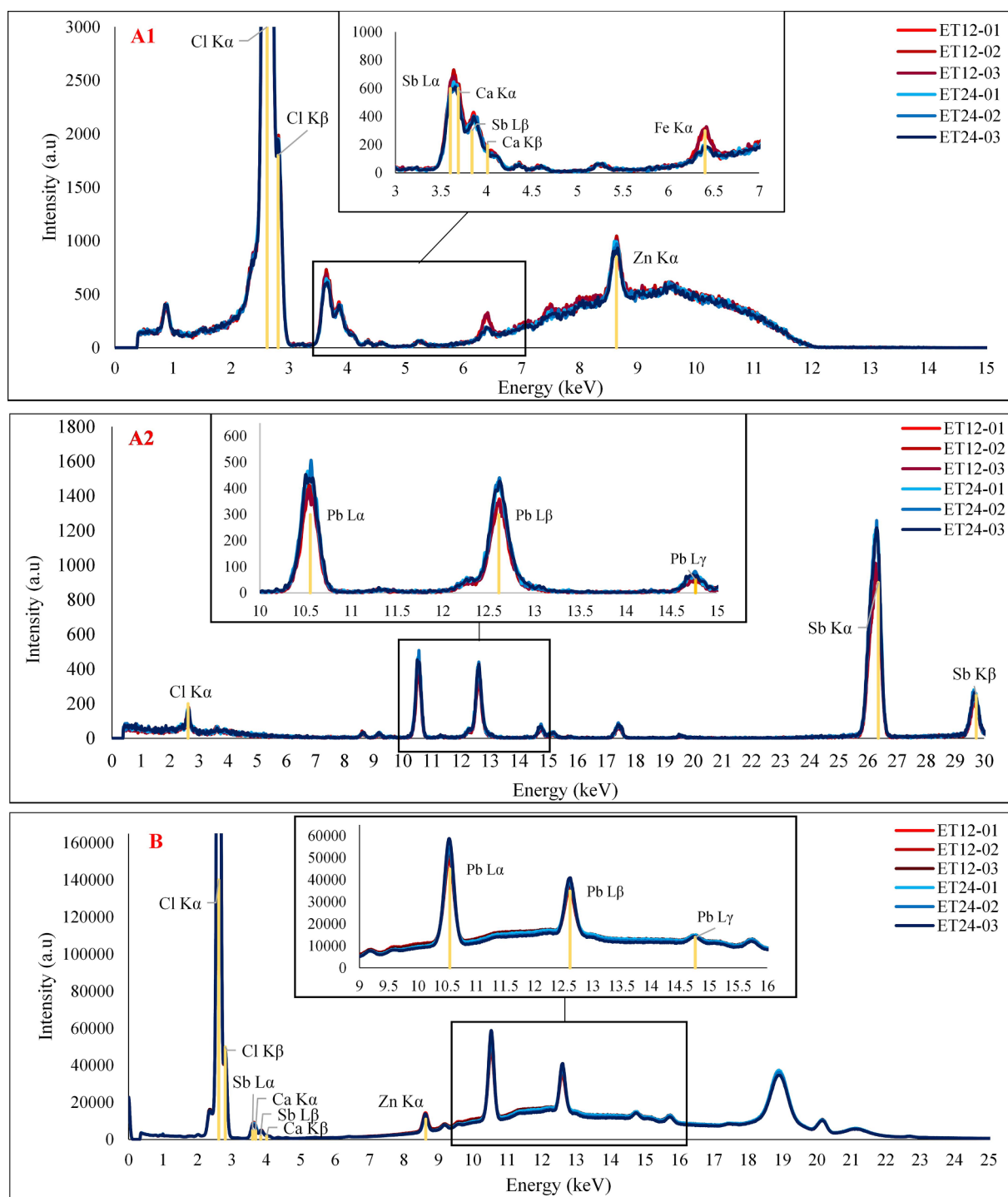
Fig. 5. Spectra overlay comparison of tapes 12 (replicates 01–03: red) and 24 (replicates 01–03: blue) for instrument B (image A1 for low filter and A2 for high filter) and instrument C (image B).

Table 6
Summary of discriminating power for the XRF instruments in this study.

	XRF A	XRF B	XRF C	LA-ICP-MS
Samples from different sources	12 (66 pairs)	12 (66 pairs)	12 (66 pairs)	12 (66)
Discrimination power (%)	84.8 (56 pairs)	97.0 (64 pairs)	90.9 (60 pairs)	90.9 (60 pairs)
Number of Distinct groups	5	10	9	9
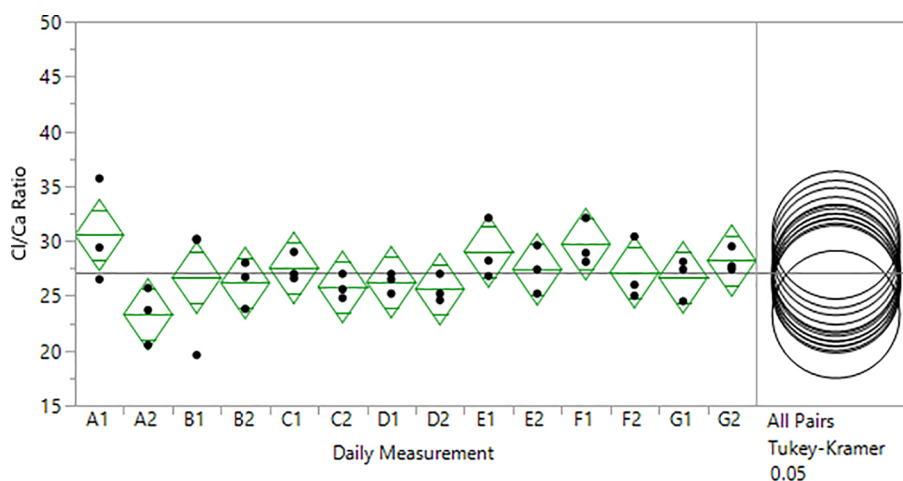
Fig 6. ANOVA and Tukey-Kramer analysis of the Cl/Ca ratio for sample 10 for the inter-day (A through G) and intra-day (1: morning, 2: afternoon) measurements, respectively.
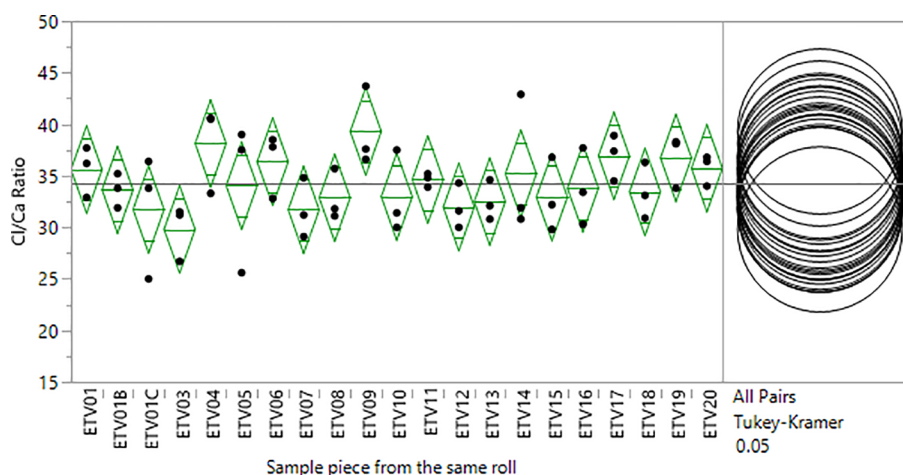
Fig 7. Analysis of the Cl/Ca ratio for the intra-roll samples by ANOVA and Tukey-Kramer.

4. Conclusions

In this study, the discrimination power was evaluated for three XRF configurations for the characterization of electrical tape backings. Forty electrical tape backings were characterized on an X-ray beam coupled to an SEM-EDS (Instrument A) and on a stand-alone XRF instrument (Instrument B). The two systems had comparable discrimination potential to LA-ICP-MS (81.5% and 90.1% vs. 84.6%, respectively). An additional standalone µ-XRF instrument (Instrument C) was used to evaluate the discrimination power between configurations. On a subset of 12 tapes, XRF C had a discrimination power of 90.9%, compared to 84.8% and 97.0% for XRF A and B, respectively. All three configurations had discrimination power greater than SEM-EDS for the subset. Inter-day, intra-day and intra-roll variation were lower than the inter-roll variations, demonstrating the utility of XRF for elemental comparisons of backings.

Utilizing XRF requires optimized sample preparation and precaution to avoid interference from the sample holders; the penetration depth of XRF can cause significant errors if the instrument measures artifacts underneath the backings. Integrating the peaks and calculating the signal-to-noise ratios for the tape samples allows for determination of the elements that are present in a sample above the limit of detection, which aids in selection of elements for use in semi-quantitative

intensity ratio calculations. Performing ANOVA and Tukey-Kramer analysis on the signal intensities aids in determining which sample pairs are significantly different and provides a more objective comparison method than spectral overlay alone.

As a result, XRF is determined to be a viable technique for the forensic examination and comparison of electrical tape backings. The study suggests that the instrumental configuration of the XRF system plays a critical role on the overall performance of the method. For instance, the stand-alone XRF system evaluated in this study provided superior capabilities in comparison to the X-ray tube attached to the SEM system. The main factors influencing the analytical performance of the XRF systems are the detector settings and the size of the beam. Since the sample size is not as limited in tape examinations as in other trace evidence materials, the examiner can afford to use collimators with large spot size to improve detection limits and to reduce acquisition times. The method offered added value to SEM-EDS elemental analysis, regarding the compositional information, classification capabilities and discrimination. Although XRF is a less sensitive and selective technique than LA-ICP-MS, among the samples tested in this study, the two techniques provided comparable discrimination power. Therefore, XRF provides an efficient, fast and cost-beneficial method for analyzing and comparing electrical tape samples.
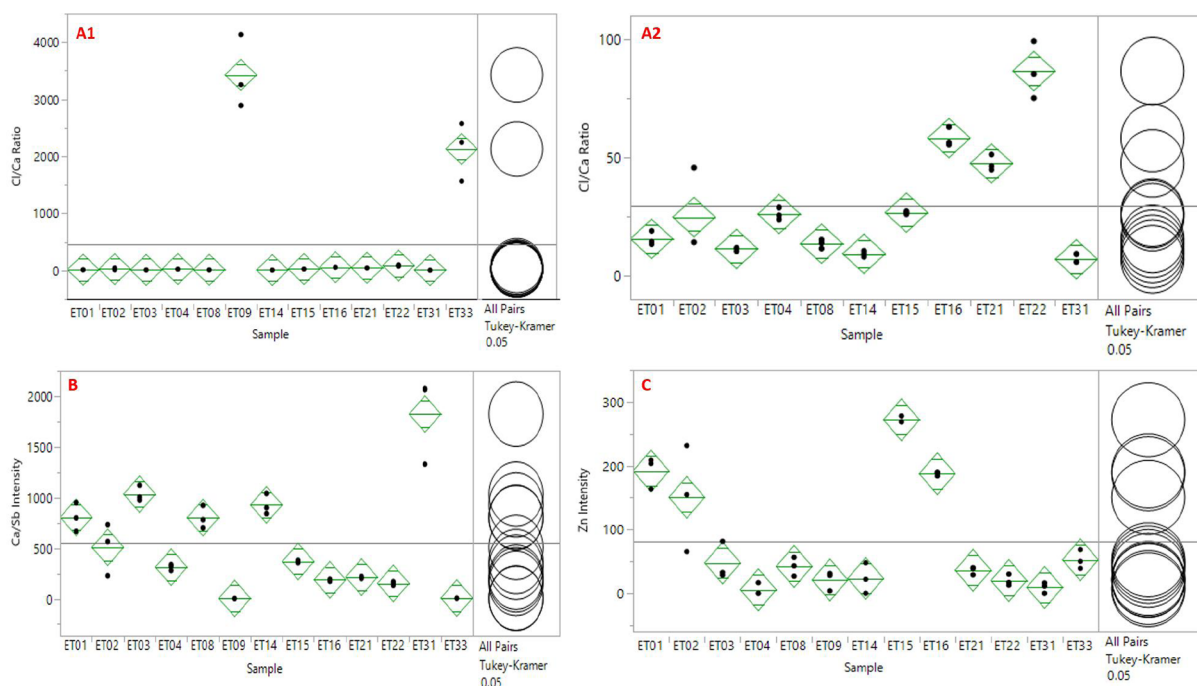
Fig 8. ANOVA and Tukey-Kramer analysis of tape samples from different sources. Image A1 demonstrates the Cl/Ca ratio analysis for some of the tape samples, while image A2 shows the variability of the Cl/Ca ratio for the same samples other than 9 and 33. Images B and C show the analysis of the signal intensities of Ca/Sb and Zn, respectively.

Acknowledgments

References

[1] Scientific Working Group on Materials Analysis (SWGMAT), Guideline for forensic examination of pressure sensitive tapes, J. Am. Soc. Trace Evidence Ex. 2 (1) (2011) 88–97.

[2] Scientific Working Group on Materials Analysis (SWGMAT), Guideline for assessing physical characteristics in forensic tape examinations, J. Am. Soc. Trace Evidence Ex. 5 (1) (2014) 34–50.

[3] R.O. Keto, Forensic characterization of black polyvinyl chloride electrical tape, Crime Lab Digest. 11 (4) (1984) 71–74.

[4] N. Okochi, M. Shirai, K. Nishijima, Pressure-sensitive adhesive tape and pressure-sensitive adhesive composition. U.S. 20060182958A, in Nitto Denko Corporation, 2006.

[5] Y. Nakagawa, S. Ishiguro, M. Natsume, S. Takada, K. Ooyama, H. Yanagida, Adhesive tape and substrate for adhesive tape. U.S. 6773806B, in Nitto Denko Corporation, 2004.

[6] P. Tucker, E. Pyun, Electrically insulating film backing. U.S. 5498476, in Minnesota Mining and Manufacturing Company, 1996.

[7] S. Fujiwara, M. Shirai, Y. Kawaguchi, Y. Nakagawa, H. Ukei, Pressure-sensitive adhesive tape or sheet. U.S 20040126542A, in Nitto Denko Corporation, 2004.

[8] A.H. Mehltretter, M.J. Bradley, D.M. Wright, Analysis and discrimination of electrical tapes: Part II Backings, J. Forensic Sci. 56 (2011) 1493–1504.

[9] Z. Sun, Y. Quan, Y. Sun, Elemental analysis of white electrical tapes by wavelength dispersive X-Ray Fluorescence Spectrometry, Forensic Sci Int. 232 (2013) 169–172.

[10] E. Margui, R. Van Grieken, X-Ray Fluorescence Spectrometry and Related Techniques: An Introduction, Momentum Press, New York, NY, 2013.

[11] T. Trejos, R. Koons, S. Becker, T. Berman, J. Buscaglia, M. Duecking, T. Eckert-Lumsdon, T. Ernst, C. Hanlon, A. Heydon, K. Mooney, R. Nelson, K. Olsson, C. Palenik, E.C. Pollock, D. Rudell, S. Ryland, A. Tarifa, M. Valadez, P. Weis, J.R. Almirall, Cross-validation and evaluation of the performance of methods for the elemental analysis of forensic glass by µ-XRF, ICP-MS, and LA-ICP-MS, Anal. Bioanal. Chem. 405 (2013) 5393–5409.

[12] C. Martinez-Lopez, T. Trejos, A.H. Mehltretter, J.R. Almirall, Elemental analysis and characterization of electrical tape backings by LA-ICP-MS, Forensic Chem. 4 (2017) 97–107.

[13] F. Kuczelinis, P. Weis, N. Bings, Forensic comparison of PVC tape backings using time resolved LA-ICP-MS analysis, Forensic Chem. (2018) in press.

[14] S. Uitdehaag, W. Wiarda, T. Donders, I. Kuiper, Forensic comparison of soil samples using nondestructive elemental analysis, J. Forensic Sci. 62 (2017) 861–868.

[15] T. Ernst, T. Berman, J. Buscaglia, T. Eckert-Lumsdon, C. Hanlon, K. Olsson, C. Palenik, S. Ryland, T. Trejos, M. Valdez, J.R. Almirall, Signal-to-noise ratios in forensic glass analysis by micro X-Ray fluorescence spectrometry, X-Ray Spectrum. 43 (2014) 13–21.

[16] T. Trejos, R. Corzo, K. Subedi, J. Almirall, Characterization of toners and inkjets by laser ablation spectrochemical methods and Scanning Electron Microscopy-Energy Dispersive X-Ray Spectroscopy, Spectrochim. Acta, Part B 92 (2014) 9–22.